A QUANTITATIVE MODEL FOR LIFESPAN CURVES

By

Matthew Witten

IMA Preprint Series # 460
October 1988

INSTITUTE FOR MATHEMATICS AND ITS APPLICATIONS
UNIVERSITY OF MINNESOTA
514 Vincent Hall
206 Church Street S.E.
Minneapolis, Minnesota 55455-0436
Phone: 612/624-6066  Fax: 612/626-7370
URL: http://www.ima.umn.edu
A QUANTITATIVE MODEL FOR LIFESPAN CURVES

MATTHEW WITTEN/ETC 02U↑

Abstract. This paper addresses the issue of a model for the total lifespan survival curve; with particular interest in the later life or geriatric years. The new model is based upon the basic premise that the population under study is a mixture of individuals comprising three major subgroups: (1) neonatal deaths, (2) standard Gompertzian-like survival, (3) geriatric survival. It is demonstrated that a standard mixture model, mixing three survival distributions, more than adequately describes survival over the entire lifespan of the population. Further, this newer model has the desirable added virtue that the model parameters may be interpreted in a biological manner.

Why Make A New Survival Model?. Survival analysis, at the demographic level, plays an important role in the projection/estimation of healthcare and health maintenance costs. Routinely, such projections, based upon lifetable analysis, are made as a means of better understanding the dynamics of age structure changes in the U.S. population. Witten(1) addressed the issue of modeling the survival distribution for neonatal deaths. The motivation for addressing this issue may be found in Figures 1–2. The curves in Figure 2 illustrate the changing dynamics of survival in the US population. Witten was able to demonstrate that, assuming a nonhomogeneous population(two different classes of death; neonatal death and non–neonatal death), the survival curve—in the neonatal years—could be made to have a more realistic dynamics. This correction required that we assume that the neonatal survival distribution was of an exponential form, while the remainder of the population(the non–neonatal group)survived according to a Gompertzian distribution.

Numerous investigators, Economos(4); E. Masoro and B.P. Yu(personal communication); and T. Johnson(personal communication) have expressed their dissatisfaction with the Gompertzian survival distribution as an adequate descriptor for the geriatric portion of the population(Many investigators prefer to use the Wiebull survival distribution or other survival distributions. This issue will not be addressed here. It is covered in Witten(5)). In general, the commonly used Gompertzian and power law survival distributions have a more rapidly decreasing geriatric tail than does the usual experimental population that they are intended to model. Issues of information content of such curves are discussed in Witten(2). While Guess and Witten(3) attempt to address the more complex question of what underlying individual mortality rates might or might not give rise to the Gompertzian–like survival distribution.

↑Computational Biology And Medicine, ETA Systems, Inc., 1450 Energy Park Drive St. Paul, MN 55108 USA

RUNNING TITLE: GERIATRIC SURVIVAL DYNAMICS
The Gompertzian survival curve is a two parameter curve based upon an exponential mortality rate function, denoted $\lambda(t)$, where

$$\lambda(t) = h_0 e^{\gamma t} \quad (1)$$

The parameters are $h_0$, the age-independent mortality rate coefficient and $\gamma$ the age-dependent mortality rate coefficient. Such a mortality rate function gives rise to the Gompertzian survival distribution $S(t)$ given by

$$S(t) = e^{\frac{h_0}{\gamma} (1 - e^{\gamma t})} \quad (2)$$

Figure 3 illustrates how $S(t)$ changes for various changes in the two mortality parameters. Figure 1 illustrates a sample comparison of the actual survival versus the Gompertzian theoretical prediction. It is quite clear that the Gompertzian tail does not adequately model the geriatric decline in the illustrated population group. Such a behavior is quite common in biological survival distribution literature.

**The Basic Hypotheses.** This paper will develop a new quantitative model for the biological survival distribution; a model which—in particular—more accurately accounts for the known dynamics of the geriatric tail. Secondly, this model should provide a description of survival dynamics over the whole lifespan of the population; from neonatal through geriatric. In order to construct this model, it is postulated that a closed, age-structured, population is comprised of individuals of three types; a neonatal failure type, a Gompertzian failure type, and a geriatric failure type (modeled by a shifted exponential distribution). How or why an individual becomes classified as a component of one of the three aforementioned groups is to be the subject of a future discussion. It is important to realize that, while we postulate the existence of these three population subgroups, statistical analysis must verify the significance of the population(mixture) proportion coefficients. Should a population proportion coefficient be statistically insignificant, then this would imply that the corresponding mixture component of the population is not significant in contributing to the survival dynamics of the given data set. For example, an insignificant neonatal coefficient implies that the neonatal component, of the population, does not contribute, in a significant manner, to the total survival dynamics of the population.

It will also be of interest to augment this model so as to be able to examine how transitions from the various population subgroups might affect the shape of the survival curve. One might wish to ask such questions as, given that all of the neonatal deaths disappeared and were equally assigned to the other components of the population, how would the survival distribution change? Such a question is relevant to discussions concerning the squaring of the human survival curve and its potential effects upon health care in the United States. For example, see the discussions of Fries(5) and Brody and Schneider(4). Hence, as the new survival model is developed, it will be generalized in such a manner so as to allow the potential for studying such transitions. The allowed transitions, between the population components, are illustrated in Figure 4. With these requirements and assumptions in mind, let me now discuss the new three-group survival model.
Formulation Of The New Survival Model. We begin the model formulation by supposing that we are working with a closed population which, at time \( t = 0 \) (start time), is composed of three types of individual; \( N_1 \) the total number of short-lived individuals, \( N_2 \) the total number of Gompertz-like individuals, and \( N_3 \) the total number of long-lived individuals (It should be noted that, for the purposes of discussion, we will use the phrase “Gompertz-like”, even though Weibull or other survival forms may be used equally well). It then follows that the total number of individuals in the population (at the start time) is given by the following equation.

\[
N_T^{(0)} = N_1^{(0)} + N_2^{(0)} + N_3^{(0)}
\]  

(1)

where \( N_T^{(0)} \) is the total number of individuals in the population at time \( t = 0 \) (indicated by the superscript \( (0) \)). In terms of the fractions of each group represented in the population, equation (1) may then be written as follows.

\[
1 = p_1^{(0)} + p_2^{(0)} + p_3^{(0)}
\]  

(2a)

where \( p_i^{(0)} = N_i^{(0)}/N_T^{(0)} \) \( i = 1, 2, 3 \) is the fraction of the \( i^{th} \) group in the population at time \( t = 0 \). Thus, if \( F_i(t) \) is the failure distribution for the \( i^{th} \) subgroup of the population, then the total failure distribution \( F(t) \) of the population is given by the weighted sum of the individual failure distributions. That is,

\[
F(t) = p_1 F_1(t) + p_2 F_2(t) + p_3 F_3(t)
\]  

(2b)

(see Witten(1) for a discussion on how one would construct equation (2b)). Remember, having determined the failure function \( F(t) \), the total survival \( S(t) \) is given by

\[
S(t) = 1 - F(t)
\]  

(2c)

It is important to be able to investigate/determine various effects of transitions, from group 1 and group 2 to group 3. That is, we wish to determine how subpopulation transitions might effect the survival curve of the total population. In order to investigate these transition effects, we shall proceed as follows.

Suppose that, at time \( t = 1 \) (which is the assumed next timestep of interest) we begin with the same total number, of individuals, or fractions of the population, as originally occurred at time \( t = 0; N_T^{(0)} \). However, now assume that this new time \( t = 1 \) population is composed of different numbers or proportions of each individual type. That is,

\[
N_T^{(1)} = N_T^{(0)} = N_1^{(1)} + N_2^{(1)} + N_3^{(1)}
\]  

(3)

where \( N_i^{(1)} \) is the number of individuals, in group \( i = 1, 2, 3 \) at time \( t = 1 \). From the transition diagram (see Figure 4), observe that the number of individuals in group 2 is
comprised of those individuals in group 2 at time \( t = 0 \), less those that transition out of group 2(\( \delta N_2^{(0)} \)), plus those individuals that transition in from group 1(\( \alpha N_1^{(0)} \)). Therefore, the new group 2 population size \( N_2^{(1)} \) is given by the following equation

\[
N_2^{(1)} = N_2^{(0)} + \alpha N_1^{(0)} - \delta N_2^{(0)}
\]  

(4a)

Following the transition diagram, and similar arguments, it is possible to observe that the number of individuals, in the new group 3, is given by

\[
N_3^{(1)} = N_3^{(0)} + \beta N_2^{(0)} + \delta N_2^{(0)}
\]  

(4b)

where the transition fractions, given in the transition diagram, are defined as follows:

\[
\delta : \text{fraction of } N_2 \text{ that become } N_3
\]

\[
\alpha : \text{fraction of } N_1 \text{ that become } N_2
\]

\[
\beta : \text{fraction of } N_1 \text{ that become } N_3
\]

It follows, from these definitions, that \( A1 - \alpha - \beta \) must be the fraction of \( N_1 \) that stay in the \( N_1 \) group of the population. Hence, we have that

\[
N_1^{(1)} = [1 - \alpha - \beta] N_1^{(0)}
\]  

(5a)

However, from equation(3), we have that

\[
1 = \frac{N_1^{(1)}}{N_T^{(0)}} + \frac{N_2^{(1)}}{N_T^{(0)}} + \frac{N_3^{(1)}}{N_T^{(0)}}
\]  

(5b)

Substitution of equations(4a,b) and equation(5a) into equation(5b) leads to the following relationship for the transition proportions.

\[
1 = \left[ (1 - \alpha - \beta) p_1^{(0)} \right] + \left[ (1 - \delta) p_2^{(0)} + \alpha p_1^{(0)} \right] + \left[ p_3^{(0)} + \beta p_1^{(0)} + \delta p_2^{(0)} \right]
\]  

(5c)

Equation(5c) follows from the fact that, as was previously mentioned, we wish to be able to examine the effects of transition(changing the proportions in each population component or survival component), upon the form or shape of the total population survival distribution. Hence, in order to do this, we will need to make use of the \( t = 0 \) values of \( N_1 \) as the baseline values and then examine what happens when transitions occur, from these baseline values, to the newer values. Within the framework of this construction, it then follows that the following statements about the new population fractions and their relationship to the old population fractions are true.

\[
p_1^{(1)} = (1 - \alpha - \beta) p_1^{(0)}
\]  

(6a)
\[ p_2^{(1)} = (1 - \delta) p_2^{(0)} + \alpha p_1^{(0)} \]  
\[ p_3^{(1)} = p_3^{(0)} + \beta p_1^{(0)} + \delta p_2^{(0)} \]  

As the population proportion (at all times) must add up to one, it follows that we can express the unknown fraction \( p_3^{(1)} \) in terms of the remaining known fractions. That is,

\[ p_3^{(1)} = 1 + (\beta - 1)p_1^{(0)} + (\delta - 1)p_2^{(0)} \]

Making use of the new fractions as described in equations (6), the total failure distribution, in terms of the new population proportions, would be expressed by the following equation.

\[ F(t) = p_1^{(1)} F_1(t) + p_2^{(1)} F_2(t) + p_3^{(1)} F_3(t) \]  

Substitution of equations (6a-c) into equation (7) leads to the following equation for the total failure distribution for our three group population.

\[ F(t) = \left[ (1 - \alpha - \beta) p_1^{(0)} \right] F_1(t) + \left[ (1 - \delta) p_2^{(0)} + \alpha p_1^{(0)} \right] F_2(t) + \left[ 1 - (\beta - 1)p_1^{(0)} + (\delta - 1)p_2^{(0)} \right] F_3(t) \]  

From equation (2c), it follows that our knowledge of the failure function, given equation (8), implies our knowledge of the associated survival function.

If we now wish to study the survival of a population, comprised of the particular proportions \( p_i^{(0)} \), simply set all of the transition fractions \( \alpha = \beta = \gamma = 0 \); thereby eliminating any effects due to transitions between the population subgroups. Then, for a given set of population proportions, one can determine the total population survival function.

Let us now compare the results of this model with the known biological results.

Comparison With Experimental Data. In the case where \( p_3 = 0 \), in the new model, the three group model is reduced to the two group model discussed, in detail, in Witten(1). Briefly, consider the early portion of Figures 1,5–6 . If, instead, \( p_3 = 0 \), we obtain what is equivalent to most laboratory experiments in which mice or rats are utilized; namely, the experimentalist waits until the neonatal portion of the population dies out (all potential neonatal deaths occur), and then he begins the experiment (see Figure 7).

Let us now illustrate a variety of comparisons between the pure Gompertz model and the new model containing both Gompertz and a geriatric term (the neonates will be ignored for the moment). Figure 8 illustrates the theoretical survival curve set for the organism \( C. elegans \) (data provided by Thomas Johnson). These theoretical curves are based upon maximum likelihood estimates for the parameters in a pure Gompertz survival model (details on maximum likelihood estimation may be found in Lee(6)). The pure Gompertz survival population is indicated by the curve labeled \( G \). The second population, the Gompertz plus geriatric population, is generated by arbitrarily setting the remaining model parameters not specified by the Gompertz population. Observe how
the addition of the geriatric component of the population smooths out the sharpness of the Gompertzian tail, making it behave more in the manner of the biological populations illustrated in Figures 1-2. The similarity of these theoretical curves to actual data is also seen by comparing Figure 8a,8b to the rotifer data (Tecane inermis) illustrated in Figure 8c. In particular, note how the addition of the theoretical “geriatric” component of the population gives rise to a dynamics well illustrated by this rotifer data set.

Now consider a simple transition model. Suppose that we begin with all of the population (millenia ago) in the neonatal component of the population. That is, suppose that all of the individuals in the population died very early on in life. Suppose that we then introduce a variety of measures which extend the lifespan of neonatal individuals such that some become geriatrics and some live in a Gompertzian manner. Finally, suppose that there are no longer any neonatal deaths. What would the new survival distribution look like? Such a scenario is illustrated by the three survival curves illustrated in Figure 9. The first curve (most lefthand curve) illustrates the pure exponential (or wild type) survival. The second curve (middle curve) illustrates what would happen if the population had been pure Gompertzian. And the third curve (most righthand curve) illustrates the full three group model. Notice how the introduction of the geriatric component of the population softens the decrease of the later age portion of the survival curve. This theoretical dynamics is well illustrated by the survival curve for the female quail (Coturnix coturnix) illustrated of Figure 9c. Observe the existence of the slight, early, neonatal dip; followed by the “Gompertzian-like” survival dynamics. Subsequently, there is the softening of the “Gompertzian” tail.

Closing Thoughts On Quantitative Survival Modeling. As was previously demonstrated, the Gompertz(or Weibull) distribution has trouble modeling populations with neonatal failure components and with long-lived components. This paper has constructed a general total survival model in which the neonatal failure and the geriatric failure, were included as two component fractions of the total population. Comparison of both the survival distribution with the known data, for a number of different survival distributions, shows that the three component population model exhibits significantly more realistic survival distributions than the Gompertzian or Gompertzian-like distributions alone.

Further, the model has the capability of describing lifespan changes over the whole range of the population lifespan, thereby making available a general model for describing the survival of a population over the complete age distribution in the population.

Acknowledgements.

I would like to thank the various investigators who have given me the freedom to make use of their survival data: T. Johnson, E. Masoro, B.P. Yu, C. Finch, J. Smith-Sonnenborn, R. Arking.

A number of individuals have contributed their time in discussing the various concepts in this paper and in prior related papers. I wish to acknowledge the contributions of the

I would like to thank Richard Greenberg for his numerous hours spent discussing the theoretical aspects of survival distribution theory, for his reading of rough drafts of this manuscript, and for his cross-checking of my computer code.

Finally, I would like to thank C.E. Finch and his aging research group for their vigorous discussion and input on various aspects of this work. I would also like to thank the referee for his careful and constructive comments.

REFERENCES


An illustration of the model based expected or average lifespan (MTTF) versus the experimental expected lifespan as calculated from various experimental *Drosophila* datasets. Data was obtained from R. Arking. Details may be found in R. Arking, *Successful selection for increased longevity in Drosophila: Analysis of the survival data and presentation of an hypothesis on the genetic regulation of longevity*, preprint 1986. Column 1 of this table indicates the strain of *Drosophila*, column 2 the generation, column 3 the experimental average lifespan, and column 4 the predicted average lifespan based upon the author's model computations.
<table>
<thead>
<tr>
<th>STRAIN</th>
<th>GENERATION</th>
<th>EXPERIMENTAL LIFESPAN (MEAN)</th>
<th>MODEL BASED LIFESPAN (MEAN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>F1</td>
<td>50.7</td>
<td>49.7</td>
</tr>
<tr>
<td></td>
<td>F5</td>
<td>58.3</td>
<td>56.97</td>
</tr>
<tr>
<td></td>
<td>F9</td>
<td>44.7</td>
<td>44.0</td>
</tr>
<tr>
<td></td>
<td>F13</td>
<td>48.0</td>
<td>47.8</td>
</tr>
<tr>
<td></td>
<td>F17</td>
<td>44.9</td>
<td>45.0</td>
</tr>
<tr>
<td></td>
<td>F22</td>
<td>49.0</td>
<td>47.89</td>
</tr>
<tr>
<td>DC-E</td>
<td>F1</td>
<td>54.7</td>
<td>54.0</td>
</tr>
<tr>
<td></td>
<td>F5</td>
<td>50.8</td>
<td>50.5</td>
</tr>
<tr>
<td></td>
<td>F9</td>
<td>34.4</td>
<td>34.8</td>
</tr>
<tr>
<td></td>
<td>F13</td>
<td>40.1</td>
<td>39.4</td>
</tr>
<tr>
<td></td>
<td>F17</td>
<td>52.9</td>
<td>53.2</td>
</tr>
<tr>
<td></td>
<td>F21</td>
<td>52.1</td>
<td>51.7</td>
</tr>
<tr>
<td></td>
<td>F26</td>
<td>48.7</td>
<td>46.4</td>
</tr>
<tr>
<td>NDC-E</td>
<td>F1</td>
<td>37.6</td>
<td>36.2</td>
</tr>
<tr>
<td></td>
<td>F5</td>
<td>40.4</td>
<td>40.2</td>
</tr>
<tr>
<td></td>
<td>F9</td>
<td>35.4</td>
<td>34.5</td>
</tr>
<tr>
<td></td>
<td>F13</td>
<td>34.4</td>
<td>34.0</td>
</tr>
<tr>
<td></td>
<td>F17</td>
<td>47.8</td>
<td>47.3</td>
</tr>
<tr>
<td></td>
<td>F21</td>
<td>52.3</td>
<td>51.7</td>
</tr>
<tr>
<td></td>
<td>F25</td>
<td>51.4</td>
<td>50.6</td>
</tr>
<tr>
<td>DC-L</td>
<td>F1</td>
<td>52.3</td>
<td>52.3</td>
</tr>
<tr>
<td></td>
<td>F5</td>
<td>38.5</td>
<td>37.5</td>
</tr>
<tr>
<td></td>
<td>F9</td>
<td>53.4</td>
<td>52.9</td>
</tr>
<tr>
<td></td>
<td>F13</td>
<td>54.7</td>
<td>54.5</td>
</tr>
<tr>
<td></td>
<td>F17</td>
<td>56.8</td>
<td>56.5</td>
</tr>
<tr>
<td></td>
<td>F21</td>
<td>59.6</td>
<td>59.3</td>
</tr>
<tr>
<td>NDC-L</td>
<td>F1</td>
<td>38.2</td>
<td>37.6</td>
</tr>
<tr>
<td></td>
<td>F5</td>
<td>37.8</td>
<td>37.0</td>
</tr>
<tr>
<td></td>
<td>F9</td>
<td>53.9</td>
<td>53.4</td>
</tr>
<tr>
<td></td>
<td>F13</td>
<td>65.7</td>
<td>65.0</td>
</tr>
<tr>
<td></td>
<td>F17</td>
<td>66.5</td>
<td>66.6</td>
</tr>
<tr>
<td></td>
<td>F21</td>
<td>71.7</td>
<td>71.5</td>
</tr>
<tr>
<td></td>
<td>F25</td>
<td>77.0</td>
<td>79.4</td>
</tr>
</tbody>
</table>
LEGENDS FOR FIGURES

[1] An illustration of Drosophila survival data from Economos(2). Notice how the theoretical and the data curves have decidedly different properties in the geriatric portion of the distribution.

[2] An illustration of the variety of survival dynamics illustrated by the U.S. survival of white males and white females during the years 1840–1960. The data is obtained from Jacobson(7). Notice the clear, early, neonatal dip and the non-Gompertzian dynamics of the right tail of the survival distribution.

[3] These figures illustrate the effect of varying the two Gompertz mortality rate parameters \( h_0 \) and \( \gamma \). In Figures 3a the parameter \( \gamma \) is fixed while \( h_0 \) is varied as follows: \( h_0 = 10^2 \), \( j = 1, 2, \ldots, 5 \). For Figure 3a.1 \( \gamma = 0.5 \), Figure 3a.2 \( \gamma = 1.0 \), and Figure 3a.3 \( \gamma = 2.0 \). Notice how, as I change the value of \( \gamma \), the “linear portion” of the Gompertz curve shifts to the left or to the right depending upon whether I increase or decrease the value of \( h_0 \). Figures 3b illustrate the effects of keeping \( h_0 \) while varying \( \gamma \). Figure 3b.1 illustrates \( h_0 = 0.1 \) for the values \( \gamma = 10^{-3} \), \( j = 1, 2, \ldots, 5 \). Figure 3b.2 illustrates \( h_0 = 0.01 \) for the same values of \( \gamma \).

[4] An illustration of the possible allowed transitions in my survival model. In particular, I allow neonates to become Gompertzian or geriatric, and I allow Gompertzians to become geriatric. The population proportions are indicated by the variables \( p_i \) \( i = 1, 2, 3 \). The transition fractions are indicated by the variables \( \alpha \), \( \beta \), and \( \delta \).

[5] Survivorship curves for 82 males and 45 females of the black widow spider Latrodectus mactans(Fabr) are illustrated in this figure.

[6] A set of lifespan curves for the pulmonate Limnaea columella at two different population densities. Data from Comfort(8). Notice the similarity of these curves to the dynamics of the full three-group population model(1) in Figures 9a, 9b.

[7] An illustration of rat survival data from the paper of Yu et al.(10). In survival experiments of this type, the neonatal component is eliminated by waiting until the neonatal deaths are zero. This is equivalent to setting \( p_2 = 0 \) in my quantitative model.

[8a] An illustration of a theoretical survival curve generated for a C. elegans dataset(taken from T. Johnson, with permission). The G curve illustrates a pure Gompertz survival curve, where \( h_0 = 0.0005842 \) and \( \gamma = 0.306 \). As based upon maximum likelihood estimates for the survival parameters. The second curve illustrates the effect of adding a geriatric component to the model. The models missing parameters have been arbitrarily set.

[8b] An illustration of a similar C. elegans dataset where the effect of the geriatric component of the population has been purposely magnified for effect.

[8c] Survivorship curves for three different types of rotifer(Taken from Comfort(8)). Observe the similarity of the A and the M curves(monic female rotifers and unfertilized motic female rotifers) to the non-Gompertzian curve illustrated in Figure 8b.

[9a] An illustration of how the capability of the three group model to adequately model the vast variety of survival dynamics found in such survival studies. Curve E is the wild or exponential survival curve in which all of the individuals in the population die very early on in life. Curve G is the pure Gompertzian survival curve which is meant to provide a standard for comparison. Finally, curve F is the full three group population model. In this figure, \( p_1 = 1 \), \( p_2 = p_3 = 0 \), \( h_0 \) and \( \gamma \) are as in Figure 8, and the transition fractions are given as \( \alpha = 0.7 \), \( \beta = 0.3 \), and \( \delta = 0.0 \). Again, notice the softening of the geriatric tail of the three group population model.

[9b] This figure is similar to Figure 9a save that \( \beta = 0.25 \). This will generate a stronger neonatal effect in the earlier population years.

[9c] A survival curve for female quail(Coturnix coturnix). The data was taken from Muller et al.(9). Observe the similarity of the quail survival dynamics to the curve labeled F in Figure 9b. Observe the early small dip in the curve, the subsequent Gompertzian-like behavior, and the final non-Gompertzian tail.