Journal of Official Statistics
Vol. 3, No. 1, 1987, pp. 19–36

Past and Recent Attempts to Model Mortality at All Ages

Michael Hartmann[1]

Abstract: In demography, model life tables have played an increasingly important role in indirect estimation, population projections, simulations and for other purposes where there is a need for a model of mortality at all ages. In actuarial science, model life tables have played a subdued role because laws of mortality, i.e., parametric functions that give a good fit to empirical mortality curves, are a better means of graduation than discrete mortality representations. Most laws of mortality are partial in the sense that they apply only to a broad age group and not to all ages. This paper focuses on three laws of mortality that apply to all ages. Two of them were developed by the actuaries Thiele and Wittstein in the late 19th century. The third, developed by Heligman and Pollard, is of recent origin. The three laws are discussed with references to Scandinavian mortality data. The results suggest that the most recently proposed law can be used for generation of model life tables, for making population projections, simulations, and other statistical work where there is a need for a realistic model of human mortality.

Key words: Laws of mortality; model life tables; indirect estimation; normal places; least squares minimization; population projections.

1. Introduction

Since the first half of the 18th century, laws of mortality, i.e., parametric functions that can be used to model empirical mortality curves, have developed into an established research topic among actuaries, demographers, and others interested in the statistical study of human mortality.

In this paper, I study three laws of mortality that are intended to model mortality at all ages. I focus on work by Thiele (1871 and 1872), Wittstein (1883) and Heligman and Pollard (1980). Because the attention is limited to mortality representations by differentiable parametric functions, traditional model life tables, i.e., tabular representations of the age pattern of mortality are not discussed.

It will appear that these three laws were developed from the same principles and, in particular, the Heligman and Pollard's law from 1980 is an enhanced version of Thiele's law from the 1870s. The results presented in this paper suggest that Heligman and Pollard's law possibly is the best existing demographic model of mortality at all ages and is an efficient means of generating model life tables, e.g., for use with population projections.

[1] Coordinator, National Census Program, Office of Planning and Statistics, Federated States of Micronesia, P.O. Box 538, Kolonia, Pohnpei, Eastern Caroline Islands 96941. The author wishes to extend his sincere thanks to Jan M. Hoem and Ingrid Lyberg for their comments on an earlier draft of this paper as well as to two anonymous referees for their constructive views. Sincere thanks are also extended to Viorica Diaconescu for drawing the graphs.

Actual generation of model life tables, however, is not attempted in this paper.

The paper begins with a presentation of the necessary statistical definitions, notation, and the three models under consideration. Section 2 focuses on estimation of the parameters, Section 3 on an application of Thiele's model, Section 4 on a comparison of Thiele's law with Wittstein's, and Section 5 is devoted to a discussion of Heligman and Pollard's model and its potential uses. This discussion is based on Swedish mortality data from the period 1900–1970. Section 6 summarizes the results of the paper.

In the discussion that follows, I use standard demographic and statistical notation, i.e., the survival function s is such that for continuous x, $s(x)$ is the probability of survival from birth until age x. The probability for a person aged x to survive to age $x + 1$ is

$$p_x = s(x+1)/s(x). \tag{1.1}$$

The probability that a person aged x dies before reaching age $x + 1$ is $q_x = 1 - p_x$, which is commonly known as the (life table) mortality rate. The force of mortality is

$$\mu(x) = -s'(x)/s(x), \tag{1.2}$$

which is also known as the mortality intensity. The intensity and the survival functions are connected by the well-known relation

$$s(x) = \exp\{-\int_0^x \mu(t)dt\}, \tag{1.3}$$

whereby

$$-\ln p_x = \int_x^{x+1} \mu(t)dt. \tag{1.4}$$

Throughout ln denotes the natural logarithmic function.

Historically, laws of mortality date back to about 1725 when de Moivre suggested that the probability of survival from birth until age x could be expressed as a linear function of age, i.e., as

$$s(x) = 1 - x/w, \tag{1.5}$$

where w is the highest attainable age. De Moivre knew that his model did not give a realistic representation of empirical survival functions, and he proposed it merely to make various actuarial calculations (Hooker and Longley-Cook (1953, p. 161)).

Perhaps the most well-known law of mortality is the one introduced by Gompertz in 1825, namely

$$\mu(x) = Bc^x \tag{1.6}$$

(see, e.g., Smith and Keyfitz (1977, pp. 279–282)). Whereas this law in many cases gives a surprisingly good fit to adult and old age mortality, it rarely gives a good fit to childhood and middle life mortality. Because there is usually little numerical difference between the estimated mortality intensity and the estimated (life table) mortality rate, (1.6) can be used to model either of the functions.

Elston (1923) has given an account of the most noteworthy laws of mortality in the 19th century actuarial literature. Most of these laws are partial in the sense that they only give good fits to sections of the mortality curve and, hence, fail in representing the mortality curve at all ages. In fact, it is a basic characteristic of laws of mortality that they fail in representing childhood and, especially, the accident hump at young adult ages.

Thiele, therefore, made actuarial and demographic history when in 1872 he published a paper on modeling mortality at all ages in the *Journal of the Institute of Actuaries (JIA)* (1872, Vol. 16, pp. 313–329). In terms of the mortality intensity, Thiele proposed the composition

$$\mu(x) = \mu_1(x) + \mu_2(x) + \mu_3(x) \tag{1.7}$$

with

$$\mu_1(x) = a_1\exp\{-b_1x\}, \tag{1.8}$$

$$\mu_2(x) = a_2\exp\{-\tfrac{1}{2}b_2(x-c)^2\}, \tag{1.9}$$

and

$$\mu_3(x) = a_3\exp(b_3x) \tag{1.10}$$

where the seven parameters a_i, b_i $(i=1,3)$ and c are positive. The hypothesis underlying Thiele's law is that the causes of death naturally fall into three classes; those affecting childhood, middle life, and adult ages. To model childhood mortality, Thiele chose the Gompertz law (1.8) which makes μ_1 vanish for adult ages. In order to capture the characteristic accident hump at young adult ages he chose (1.9) which, apart from a scale factor, is the normal probability density function. It will be seen that μ_2 is likely to give small values for childhood and old ages. Finally, adult and old age mortality is modeled by the Gompertz law (1.10) which ensures that μ_3 vanishes for childhood and middle life ages. Jointly, the three functions give a model of mortality at all ages.

Even though Thiele was a professor of astronomy in Copenhagen from 1875 to 1907, he made lasting contributions to mathematical statistics and actuarial science. Regrettably some of his most important actuarial work was never published, but he did publish his laws of mortality which became well-known among contemporary actuaries (Hoem (1983, pp. 215–217)).

About 10 years after Thiele had published his law (1.7) in the *JIA*, Wittstein (1883) published in the same journal his law

$$q_x = \frac{1}{m}a^{-(mx)^n} + a^{-(M-x)^n}, \tag{1.11}$$

which models q_x between ages 0 and M, where M is the highest attainable age. It may have been Wittstein's hypothesis that the causes of death naturally divide into childhood and adult age components.

Wittstein (1883) reached (1.11) on the basis of an elegant generalization of de Moivre's law (1.5), and mentions that he is the first to consider the problem of constructing a (realistic) law of mortality for all ages. Evidently Wittstein cannot have been familiar with Thiele's work that was published in the *JIA* about ten years earlier. A language barrier, perhaps, accounts for this circumstance. Thiele's paper in *JIA* was a translation made by Sprague from an original paper written in Danish (Thiele (1871)). By the same token, Wittstein's paper in *JIA* was a translation made by Bumsted from the original paper written in German.

Wittstein's law appears to have received somewhat more attention in the literature than Thiele's (see e.g., Henderson (1915, pp. 31–32), Wolfenden (1942, p. 85)). As a curiosity, I also mention that Wittstein's law has been used by Statistics Sweden for smoothing the death rates at advanced ages.

Recently, Heligman and Pollard (1980) have proposed the eight-parameter model

$$q_x / p_x = A^{(x+B)^C} + D\exp\{-E(\ln x - \ln F)^2\} + G H^x, \tag{1.12}$$

with positive parameters $A, \ldots, H$. Because (1.12) is defined for $x > 0$ only, it requires a slight modification to define it at age 0. A study of Heligman and Pollard's paper (1980) suggests that at age 0 they let

$$q_0 / p_0 = A^{B^C} + G, \tag{1.13}$$

which one will see is $\lim_{x\to 0}(q_x / p_x)$.

With the modification (1.13), (1.12) has been applied to a range of Australian mortality experiences with promising results (Heligman and Pollard (1980)). In a working paper (Hartmann (1983)) of which the present paper is a condensed version, it is demonstrated that (1.12) gives an equally good representation of Swedish mortality between 1900 and 1970.

It is now appropriate to turn to a discussion of how to estimate the parameters in the models under discussion.

2. Fitting a Law of Mortality

2.1. *Estimating the mortality rate and the mortality intensity*

If E_x is the number of persons exposed to risk at exact age x, and if D_x is the corresponding number of persons who die before reaching age $x + 1$, the mortality rate is estimated by

$$\hat{q}_x = D_x / E_x. \tag{2.1}$$

In practice, when vital registration data are used, E_x cannot always be directly observed. Nevertheless, with ordinary observational plans E_x can be accurately estimated. In the sequel, it is assumed that E_x and D_x are accurate quantities. The variance of $\hat{q}_x$ can be estimated by

$$\hat{\sigma}^2(\hat{q}_x) = \hat{p}_x\hat{q}_x / E_x, \tag{2.2}$$

where $\hat{p}_x = 1 - \hat{q}_x$.

Assuming that the mortality intensity is piecewise constant, it is estimated by

$$\hat{\mu}(x) = -\ln \hat{p}_x, \tag{2.3}$$

which is in accordance with the result of (1.4). It is also made clear from (2.2) and (2.3) and by using a standard approximation (see e.g., Cramer (1963)) that an estimate of the variance of $\hat{\mu}(x)$ is

$$\hat{\sigma}^2(\hat{\mu}(x)) = \hat{q}_x / \hat{p}_x E_x. \tag{2.4}$$

2.2. *The method of least squares*

If $\mu(x; \mathbf{a})$, with parameter vector $\mathbf{a} = (a_1, \ldots, a_r)$, is a model of the mortality intensity, and estimated intensities $\mu(x_i)$ for the ages $x_1,\ldots,x_N$ are given, the model is fitted to these estimates by the method of least squares by minimizing

$$\sum_x w_x\{\hat{\mu}(x) - \mu(x; \mathbf{a})\}^2 \tag{2.5}$$

with respect to $\mathbf{a}$. Here w_x is a weight that applies to the single year age group beginning at age x and which should preferably be proportional to the reciprocal of the estimated variance (2.4).

Having estimated $\mathbf{a}$ by $\hat{\mathbf{a}}$, one represents the intensity function by

$$\mu^*(x) = \mu(x; \hat{\mathbf{a}}).$$

Analogically, one fits a model of mortality rates $q(x; \mathbf{a})$ by $q^*(x) = q(x; \hat{\mathbf{a}})$.

Clearly, if $\mu(x;\mathbf{a})$ is given by (1.7), minimizing (2.5) gives nonlinear normal equations that are difficult to solve unless one makes use of a computer and a sub-routine for fitting nonlinear least squares.

Although Thiele (1872, p. 319) was known as a computational wizard (Hoem (1983, p. 217)), he too found it prohibitively laborious to estimate the parameters in his law by the method of least squares. Actually, according to Jörgensen (1913, p. 116) "Thiele postulierte gewöhnlich, dass der grösste Vorteil, welchen man durch Studium der Methode der Kleinsten Quadrate erreichen könnte, darin bestunde, dass man dadurch entdeckte, wie man die Verwendung vermeiden könnte."[2]

[2] In the present author's translation: Thiele often stated that the advantage of applying the method of least-squares is that it gives to hand why it should be rejected.

2.3. *The method of normal places*

Thiele favored the method of normal places of which he has given a detailed account in his *Theory of Observations* (1903, pp. 105–110). To explain this method, let $\mu(x;\mathbf{a})$, $\mathbf{a} = (a_1,\ldots,a_r)$, be a model of the mortality intensity. Furthermore, let $\hat{\mu}(x_i)$ be estimates of the piecewise constant intensities at ages $x_1,\ldots, x_N$ where $N \geqslant r$. These can be grouped into r adjacent groups, i.e., into as many groups as there are parameters. For each of these groups one may calculate a running average. Let t_k denote the running average of the observed intensities in group k and z_k, the age to which t_k corresponds. The r points (z_k, t_k) are called normal places. The normal places, then, are smoothed intensities and the ages to which they correspond.

To estimate $\mathbf{a}$, one solves the r equations

$$t_k - \mu(z_k;\mathbf{a}) = 0;\ k = 1, \ldots , r;$$

with respect to $\mathbf{a}$.

The idea behind this method is that the application of a running average to the observed statistic produces a more accurate statistic. If one creates as many normal places as there are parameters in the model, the fitted model will agree with these normal places. Consequently, if the model is well suited for its purpose, it provides a means of reliable interpolation between normal places.

Here it is in place to digress for a moment and note that the method of normal places, which originates in astronomy (Thiele (1903, p. 106)), certainly is commonly used. However, it seems no longer to answer to the name once given to it. Those familiar with the Brass logit model will see that it is a simple version of the method of normal places which Brass recommends for fitting his model to an observed survival function (Brass (1971, p. 84)).

3. Fitting Thiele's Law to a Male Experience from About 1870

To illustrate his law, Thiele chose a set of data collected by Oppermann who was the founding father of actuarial science in Denmark (Hoem (1983, p. 215), Thiele (1871 and 1872)). This experience is based on 48 470 life-insured male Danes between the ages 5 and 85 years, giving E_x and D_x, so that the mortality rates and the intensities are readily estimated from (2.1) and (2.3).

For reasons already given, Thiele used a normal place method to estimate the parameters. To see how close he came to achieving an optimal fit in the sense of weighted least squares, I have minimized (2.5) with $\mu(x;\mathbf{a})$ given by (1.7) and the weight w_x given by the reciprocal of (2.4) for the ages 5,...,23, 26, 28,...,79. (Because $D_x = 0$ for $x = 24$, 25 and 27 for the data used by Thiele, these ages have been excluded when minimizing the sum of weighted least squares.) The minimization was accomplished using the NAG-Library routine EO4FDF which does not require explicitly given partial derivatives.

The squared standardized residual

$$R_x = [D_x - E_x q(x;\hat{\mathbf{a}})]^2 / E_x p(x;\hat{\mathbf{a}}) q(x;\hat{\mathbf{a}}),$$

with $q(x;\hat{\mathbf{a}}) = 1 - \exp\{-\mu(x+0.5;\hat{\mathbf{a}})\}$, indicates the goodness of fit. For the least squares fit, the sum of squared standardized residuals for the above ages is asymptotically Chi-square distributed with 65 degrees of freedom because there are 72 observed (asymptotically) independent and normally distributed intensities and seven parameters.

Using $\mu^*(x)^T$ and $\mu^*(x)^L$ to denote the fitted intensity given by Thiele and the one obtained by the method of weighted least squares respectively, I have shown (Fig. 1) the curves resulting from plotting $10\ \ln(10^5\mu^*(x)^T)$, $10\ln(10^5\mu^*(x)^L)$ and $10\ln(10^5\hat{\mu}(x))$ against age x. This transformation of the intensity function is particularly useful when one wishes to inspect to quality of the fit. The estimated parameters for $\mu^*(x)^T$ and $\mu^*(x)^L$ as well as the sum of squared residuals for the two fits are given in Table 1.

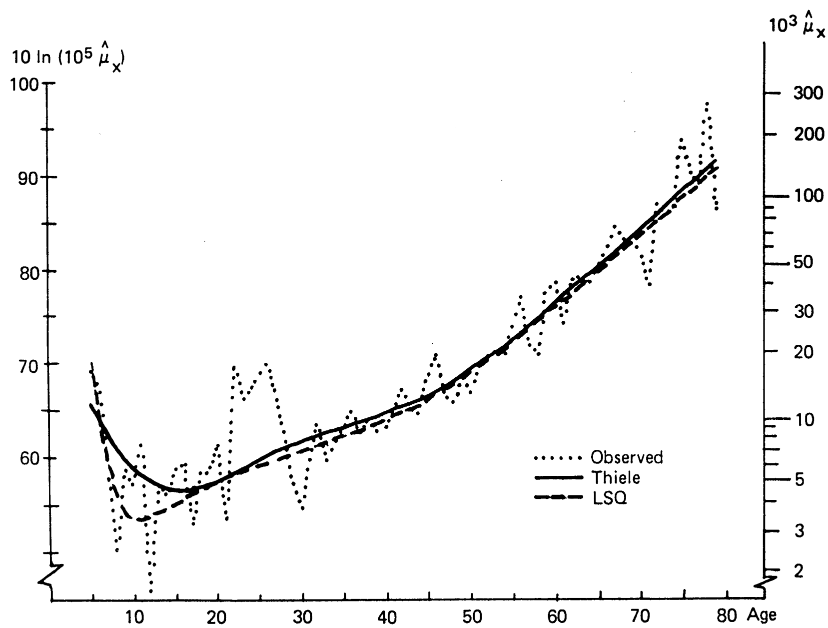


Fig. 1. Observed and fitted mortality intensities using a Danish late 19th century experience and Thiele's law of mortality

Table 1. Estimated parameters and goodness of fit for Thiele's law of mortality

Method	a_1	b_1	a_2	b_2	c	a_3	b_3
Normal places	0.02474	0.17028	0.00483	0.08238	31.709	0.0002699	0.07990[1]
Weighted least squares	0.68689	0.75540	0.00413	0.05898	27.197	0.0003140	0.07683

	Goodness of fit[2]			
	All ages	5–19	20–39	40–79
Normal places	76.68	26.43	14.87	35.38
Weighted least squares	69.33	16.98	14.95	37.40

[1] In Thiele's paper in the *JIA* (1872, p. 323), but not in the original Danish paper (Thiele (1871)), this parameter is given incorrectly as 0.7990.
[2] Sum of squared standardized residuals; $\chi^2_{72-7,\,0.05} = 84.8$.

It will be seen (Fig. 1) that relative to the normal place method, the method of weighted least squares gives a substantially closer fit for childhood ages. On the other hand, it will also be seen that for middle life and adult ages, Thiele's fit is slightly better than the least squares one. From a practical point of view, Thiele was probably more interested in knowing if his formula gave a good fit to ages of actuarial interest than to childhood ages. If so, it must be concluded that Thiele, for all practical purposes, achieved something very close to an optimal fit to the Danish experience.

It should be noted that for both fits the total sum of squared residuals is well below the critical value for a test on the 5 percent level, say, which is $\chi^2_{.05} = 84.8$. This suggests that both fits could be seen as a graduation of the experience. That is, the fitted deaths are statistically commensurate with the observed ones. In further support of this, it will be seen (Fig. 1) that the sign changes between observed and fitted intensities vary in a non-systematic manner. In effect, Thiele's application of his law to the Danish experience was a successful actuarial experiment.

4. Comparing the Laws of Thiele and Wittstein

Since Wittstein's law (1.11) does not contain a term for middle life mortality, it cannot model the accident hump. To show how it models the mortality curve *in toto,* it has been fitted to the Swedish female experience for the period 1961–1970 by means of the (unweighted) method of least squares (Fig. 2). The minimization was done for the ages 1, ... , 75 using the previously mentioned NAG routine. The estimated parameters for Wittstein's law are $\hat{m} = 22.6039$, $\hat{a} = 1.8731$, $\hat{n} = 0.6043$ and $\hat{M} = 88.6056$.

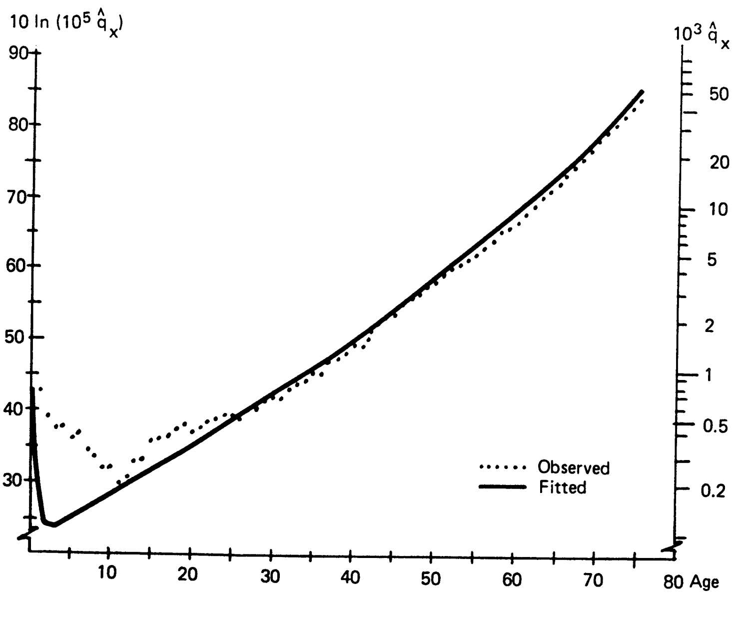


Fig. 2. Observed and fitted mortality rates using the Swedish female experience for 1961–70 and Wittstein's law of mortality

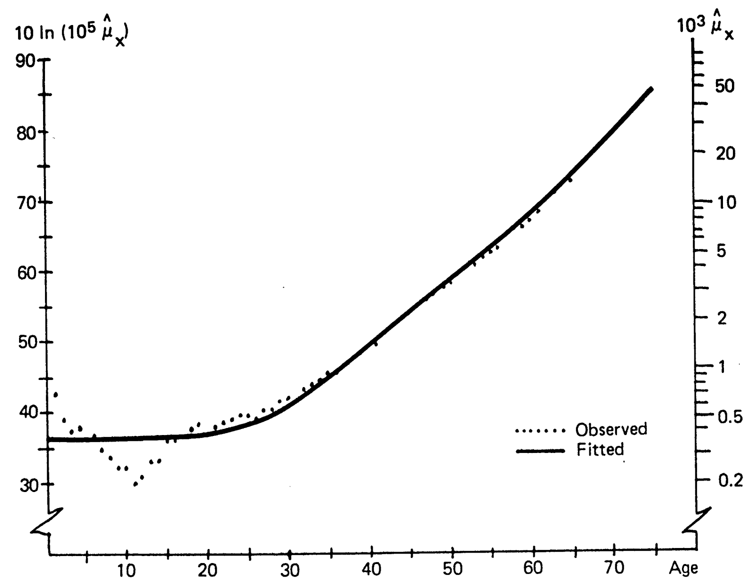


Fig. 3. Observed and fitted mortality intensities using the Swedish female experience for 1961–70 and Thiele's law of mortality

Applying Thiele's law (2.5) to the same experience and the analogical (unweighted) method of least squares (Fig. 3), the estimated parameters become $\hat{a}_1 = 0.00307$, $\hat{b}_1 = 0.00493$, $\hat{a}_2 = 0.00129$, $\hat{b}_2 = 0.0904$, $\hat{c} = 53.801$, $\hat{a}_3 = 0.00307$ and $\hat{b}_3 = 0.129$[3].

Comparing Figs. 2 and 3 we see that although both fits are somewhat mediocre, Thiele's law provides a closer fit to the experience than Wittstein's law. Generally speaking, the superiority of Thiele's trident law over Wittstein's is obviously that it contains a ligament that binds childhood and adult mortality by the means of a middle life mortality component. It is the middle life mortality component that can reproduce the accident hump. However, when the method of least squares is used to fit Thiele's law it occasionally happens that the middle life component (1.9) is fitted to the wrong part of the mortality curve. This is seen in the case of Fig. 3 ($\hat{c} = 53.801$) which yields a "false" accident hump situated much above the real accident hump in the underlying experience. (Clearly, when this happens, the fit to adult mortality is very close.) In my computations (Hartmann (1983)), this happened when there was a more or less pronounced concavity in the adult part of the mortality curve (Fig. 3). This, of course, would not happen if one used a normal place method of estimation. The normal place method predetermines the placement of the center of the accident hump in the fitted curve. Thiele's dislike for the method of least squares is well justified by these results.

In conclusion, we see that Wittstein's law, even though it is based on an ingenious derivation of de Moivre's law (1.5), does not compete well with Thiele's dynamic mortality model.

[3] Because the numerical difference is marginal between estimates of q_x and $\hat{\mu}_x$, the two fitted curves are very close to each other. For this reason the results have been given in separate diagrams.

Interestingly enough, Heligman and Pollard's law may also result in a false accident hump situated much above the location of the real accident hump (Heligman and Pollard (1980, p. 59)). Computations made using both models (Hartmann (1983)) suggest, however, that Heligman and Pollard's law is much less likely to produce a false accident hump than Thiele's law. Consequently, it would seem that the "lognormal" nature of their middle life mortality component fits empirical experiences better than the "normal density" component used by Thiele. I now turn to a discussion of Heligman and Pollard's law.

5. Heligman and Pollard's Law of Mortality

5.1. *Its source and origin*

It would appear that Thiele's law (1.7) is the *fons et origo* of (1.12). The *modus operandi* is the same, i.e., childhood, middle life, and adult mortality are modeled separately and then put together in a joint expression. However, rather than relying on the Gompertz law to model childhood mortality, a more complicated expression

$$q_x/p_x = A^{(x+B)^C}, \tag{5.1}$$

is used. To model the accident hump, a lognormal expression is used instead of the normal one used by Thiele. In the case of adult and old age mortality, both laws make use of the Gompertz law.

5.2. *Estimation of the parameters*

Letting $Q(x,\mathbf{e})$ denote the right-hand side of (1.12), it is reasonable to expect that the parameters can be estimated by minimizing the (unweighted) sum of squares

$$\sum_x (\hat{q}_x/\hat{p}_x - Q(x;\mathbf{e}))^2,$$

with respect to $\mathbf{e} = (A, \ldots, H)$, for some range of ages $x_1, \ldots, x_m$ where x_1 is 0 or 1 and x_m about 75 years. Yet, computations made using Swedish mortality data for the seven decades between 1900 and 1970 frequently resulted in negative estimates of B, (Hartmann (1983)) which is not permitted.

To estimate the parameters, it is important to follow the procedure given by Heligman and Pollard (1980, p. 51), i.e., to minimize the sum of squares

$$\sum_x (q(x;\mathbf{e})/\hat{q}_x - 1.0)^2, \tag{5.2}$$

with respect to $\mathbf{e}$. Here $q(x;\mathbf{e})$ is the model rate given by (1.12). With this procedure, acceptable estimates of $\mathbf{e}$ were always obtained.

5.3. *Goodness of fit*

If one fits (1.12) to an empirical experience for which the exact number of person-years is known, one can compute squared standardized residuals so that one can see if the recorded deaths are statistically commensurate with the ones estimated from the model. When the number of person-years is very large, as in the case of the Swedish life tables, most model curves deviate significantly (in a statistical sense) from the observed curves because the variances of the estimated intensities are virtually zero. Despite this, a visual inspection of the fit may convince one that the fit is "very good" or, at least, "good enough" for practical work.

Heligman and Pollard's law gives a close fit to the 14 Swedish life tables for males and females between 1900 and 1970. Because these experiences and graphs have been given elsewhere (Hartmann (1983)), and because the graphs shown in this paper are representative, in terms of goodness of fit, of all the Swedish life tables, it is sufficient to limit the discussion to four graphs.

Figs. 4 and 5 show the female experiences for the periods 1901–10 and 1941–50 and

Figs. 6 and 7 the male experiences for the periods 1921–30 and 1961–70. Hence, the four figures give four entirely different age patterns of mortality.

Here, 10 $\ln(10^5 q^*(x;\hat{\mathbf{e}}))$ and 10 $\ln(10^5\ \hat{q}_x)$ have been shown because this transformation is particularly useful for inspecting the goodness of fit to childhood mortality and to the accident hump.

The parameters were estimated by minimizing the (unweighted) sum of squares (5.2) for the ages 0, ... , 75. Tables 2 and 3 give the estimated parameters and their correlations. The reason for studying the correlations between

Table 2. Estimated parameters in Heligman and Pollard's law using Swedish life tables for the period 1901–1970

Sex and Period	Estimated Parameters							
	A	*B*	*C*	*D*	$E'=1/E$[1]	*F*	*G*	*H*
Males								
1901–10	0.0208178	0.0682159	0.197297	0.004161	4.64780	24.5130	0.0002788	1.07592
1911–20	0.0151257	0.0444645	0.167090	0.006081	4.71095	24.9939	0.0001853	1.08234
1921–30	0.0086684	0.0230977	0.152850	0.003322	4.49795	24.0662	0.0001408	1.08556
1931–40	0.0054870	0.0126087	0.131521	0.002191	5.45556	23.0064	0.0001268	1.08735
1941–50	0.0029210	0.0063421	0.102907	0.001591	8.20958	22.7692	0.0000716	1.09507
1951–60	0.0017749	0.0065336	0.097249	0.000763	7.59628	22.1301	0.0000388	1.10447
1961–70	0.0010330	0.0014763	0.076796	0.000673	8.29704	22.1441	0.0000359	1.10553
Females								
1901–10	0.0213801	0.115998	0.200894	0.005283	1.18497	32.0410	0.0000237	1.11154
1911–20	0.0153350	0.080996	0.169448	0.005614	1.89881	28.3678	0.0000364	1.10537
1921–30	0.0084117	0.044496	0.157024	0.003326	2.09221	26.4537	0.0000628	1.09639
1931–40	0.0049230	0.025717	0.137691	0.002155	3.18337	24.4423	0.0000768	1.09397
1941–50	0.0025565	0.019385	0.118634	0.001143	3.52734	25.4410	0.0000392	1.10239
1951–60	0.0014459	0.015960	0.109559	0.000315	2.09850	30.1329	0.0000190	1.11172
1961–70	0.0007562	0.001956	0.076678	0.000164	6.68444	21.7992	0.0000236	1.10446

Source: Life tables for the periods 1901–10, 1911–20, 1921–30, 1931–40, 1941–50, 1951–60, and 1961–70. Statistics Sweden, Stockholm, Sweden.

[1] Notice that it is $E'=1/E$ that gives the dispersion of the accident hump in (1.12).

Table 3. Correlations between parameters in Heligman and Pollard's law using Swedish life tables for the period 1901–1970

	A	*B*	*C*	*D*	$E'=1/E$	*F*	*G*	*H*
Males								
A	1.000	0.995	0.963	0.865	–0.805	0.909	0.980	–0.906
B		1.000	0.942	0.819	–0.758	0.866	0.967	–0.866
C			1.000	0.865	–0.912	0.926	0.977	–0.967
D				1.000	–0.828	0.979	0.818	–0.852
E'					1.000	–0.876	–0.839	0.896
F						1.000	0.882	–0.912
G							1.000	–0.954
H								1.000
Females								
A	1.000	0.996	0.937	0.950	–0.665	0.674	–0.085	0.291
B		1.000	0.942	0.935	–0.709	0.732	–0.125	0.344
C			1.000	0.934	–0.831	0.704	0.166	0.095
D				1.000	–0.675	0.555	0.118	0.074
E'					1.000	–0.872	–0.093	–0.219
F						1.000	–0.378	0.657
G							1.000	–0.938
H								1.000

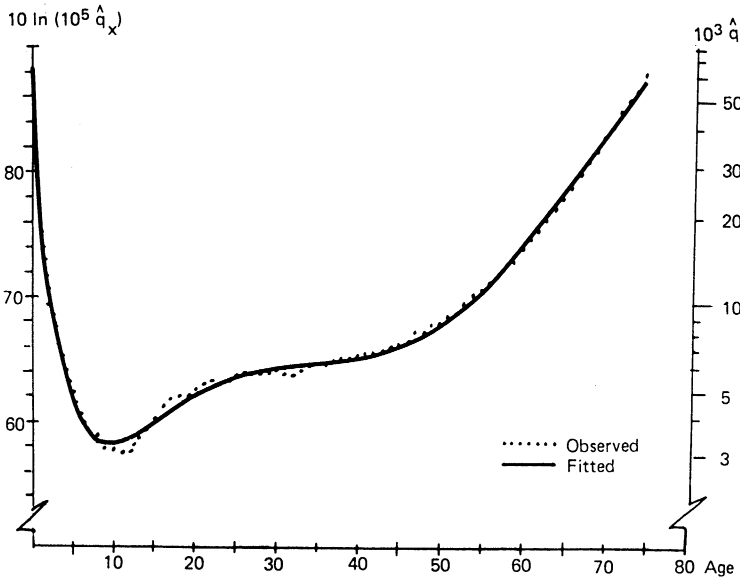


Fig. 4. Observed and fitted mortality rates using the Swedish female experience for 1901–10 and Heligman and Pollard's law of mortality

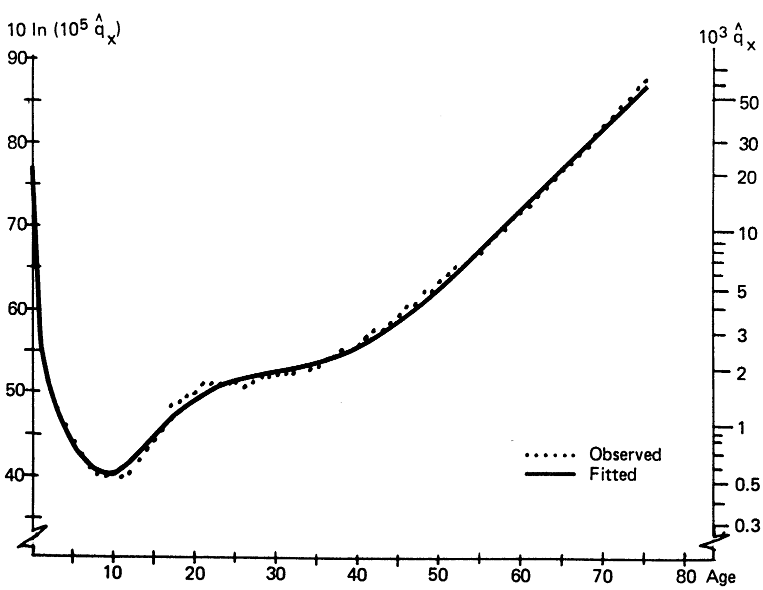


Fig. 5. Observed and fitted mortality rates using the Swedish female experience for 1941–50 and Heligman and Pollard's law of mortality

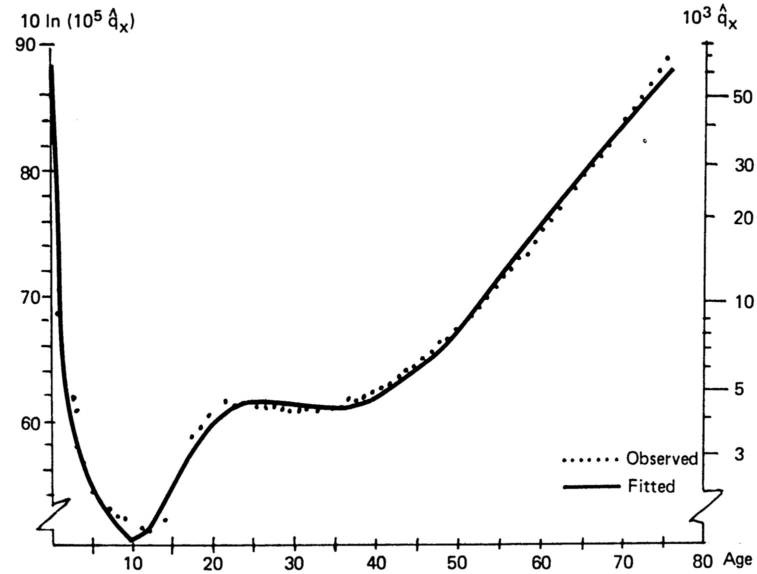


Fig. 6. Observed and fitted mortality rates using the Swedish male experience for 1921–30 and Heligman and Pollard's law of mortality

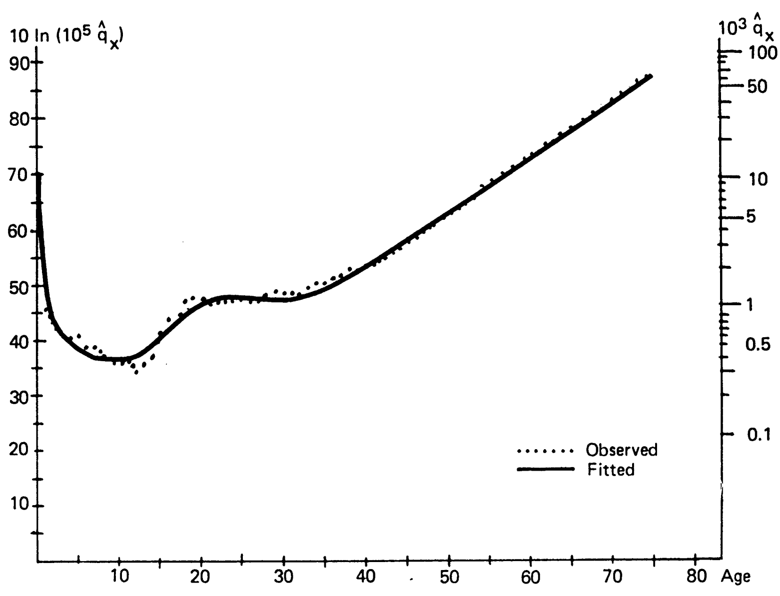


Fig. 7. Observed and fitted mortality rates using the Swedish male experience for 1961–70 and Heligman and Pollard's law of mortality

the parameters will become clear as we proceed.

It is quite obvious (see Figs. 4–7) that no existing model life table system could give equally good fits. It is, in fact, surprising that even though (1.12) has eight parameters, the modeled curve virtually coincides with the estimated probabilities between birth and age 75. I now turn to a discussion of how to interpret the parameters in (1.12).

5.4. Interpretation of the childhood parameters

Heligman and Pollard (1980) have interpreted the parameters in (1.12) to the effect that A nearly equals q_1, B is a measure of the difference between q_0 and q_1, and C is a measure of how rapidly the child adjusts to the environment. Parameter D is a measure of the level of middle life mortality, and E and F indicate the dispersion and location, respectively, of the accident hump. Finally, G and H are measures of the level, and increase with age, respectively, of old age mortality. Because the interpretations of the parameters D, E, F, G and H are fairly straightforward, the main focus is on the childhood parameters A, B and C.

For childhood ages, it is the partial law (5.1) that determines the quality of the fit given by (1.12) to empirical experiences. Consequently, if one wishes to study the behavior of the parameters A, B and C in (1.12), it is sufficient to fit (5.1) to a range of childhood experiences. Here it is appropriate to note that there have been relatively few attempts to model childhood mortality. For this reason it is interesting in its own right to fit (5.1) to childhood experiences. Estimates of A, B and C when fitting (5.1), by means of (5.2), to the abovementioned Swedish data, for the ages 0, ... , 10, are given in Table 4. To see how these estimates relate to one another, the correlations between A, B, C, $\hat{q}_0$, $\hat{q}_1$, $\hat{q}_1/\hat{q}_0$ and $\hat{q}_0-\hat{q}_1$ have been given in Table 5.

Table 4. Estimated parameters for the childhood component in Heligman and Pollard's law using Swedish life tables for the period 1901–1970

Sex and Period	Estimated Parameters						
	A	B	C	q_0	q_1	q_1/q_0	q_0-q_1
Males							
1901–10	0.019128	0.039499	0.17101	0.09255	0.02277	0.24603	0.06978
1911–20	0.014047	0.024262	0.14491	0.07643	0.01734	0.22687	0.05909
1921–30	0.007891	0.009406	0.12766	0.06472	0.01139	0.17599	0.05333
1931–40	0.005130	0.004496	0.10891	0.05080	0.00679	0.13366	0.04401
1941–50	0.002875	0.002745	0.08888	0.03034	0.00298	0.09822	0.02736
1951–60	0.001666	0.001369	0.07581	0.02022	0.00163	0.08061	0.01859
1961–70	0.000980	0.000079	0.05383	0.01529	0.00095	0.06213	0.01434
Females							
1901–10	0.018098	0.053980	0.16322	0.07598	0.02121	0.27915	0.05477
1911–20	0.013506	0.040001	0.14181	0.06112	0.01598	0.26145	0.04514
1921–30	0.007137	0.014142	0.12269	0.05052	0.00969	0.19181	0.04083
1931–40	0.004313	0.007402	0.10782	0.03878	0.00565	0.14569	0.03313
1941–50	0.002409	0.008551	0.10049	0.02329	0.00240	0.10305	0.02089
1951–60	0.001375	0.007667	0.09474	0.01546	0.00140	0.09056	0.01406
1961–70	0.000721	0.000589	0.06524	0.01151	0.00073	0.06342	0.01078

Source: Life tables for the periods 1901–1910, 1911–1920, 1921–1930, 1931–1940, 1941–1950, 1951–1960, and 1961–1970. Statistics Sweden, Stockholm, Sweden.

Table 5. Correlations between parameters for the childhood component in Heligman and Pollard's law using Swedish life tables for the period 1901–1970

	A	B	C	q_0	q_1	q_1/q_0	q_0-q_1
Males							
A	1.000	0.986	0.959	0.960	0.995	0.973	0.936
B		1.000	0.910	0.905	0.967	0.922	0.870
C			1.000	0.991	0.972	0.985	0.988
q_0				1.000	0.980	0.991	0.997
q_1					1.000	0.987	0.961
q_1/q_0						1.000	0.981
q_0-q_1							1.000
Females							
A	1.000	0.982	0.950	0.965	0.997	0.974	0.936
B		1.000	0.928	0.906	0.967	0.932	0.864
C			1.000	0.966	0.955	0.968	0.958
q_0				1.000	0.979	0.987	0.995
q_1					1.000	0.984	0.955
q_1/q_0						1.000	0.975
q_0-q_1							1.000

An inspection of the correlations (Table 5) reveals that they are all very high. In order to study the functional relationships between the parameters, I have plotted A against $\hat{q}_1$, and B and C against $\hat{q}_0$ (Figs. 8, 9 and 10).

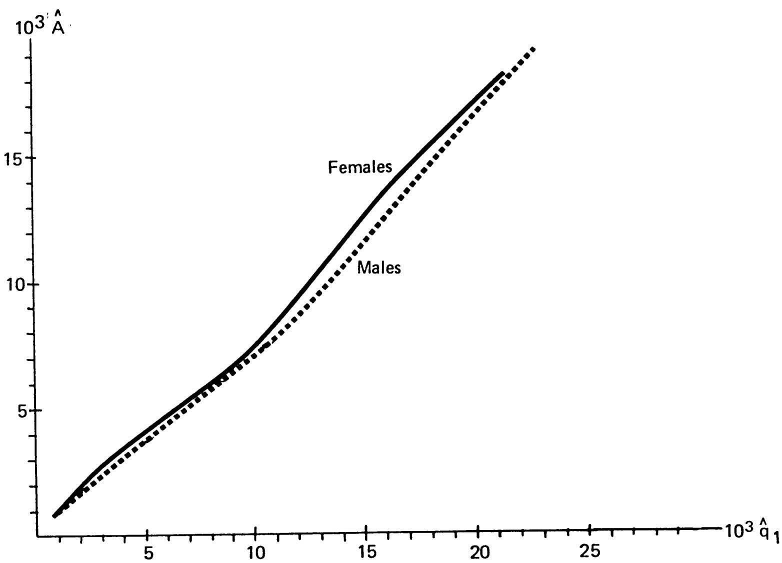


Fig. 8. The empirical relationship between parameter A and q_1 in the childhood component of Heligman and Pollard's law of mortality using Swedish life tables for the period 1901–70

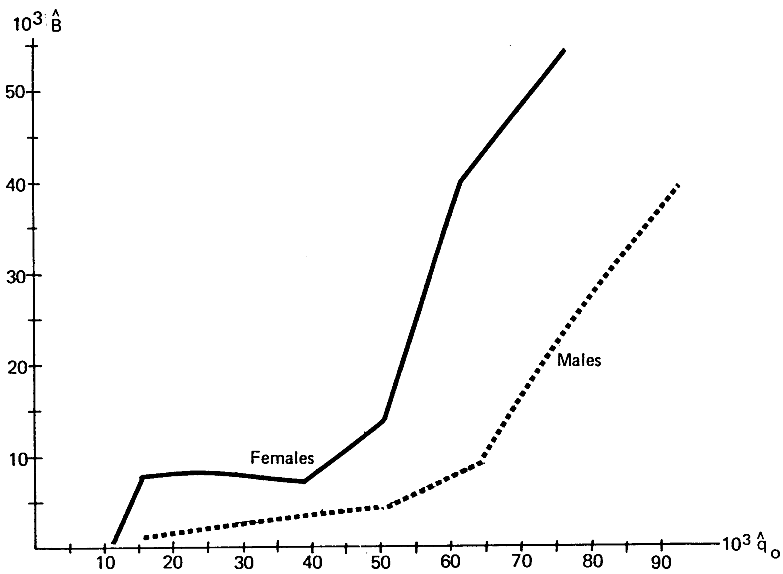


Fig. 9. The empirical relationship between parameter B and q_0 in the childhood component of Heligman and Pollard's law of mortality using Swedish life tables for the period 1901–70

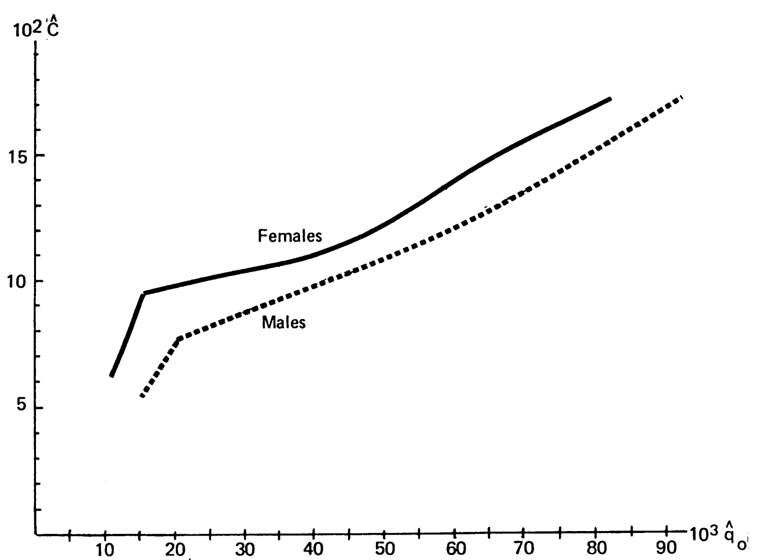


Fig. 10. The empirical relationship between parameter C and q_0 in the childhood component of Heligman and Pollard's law of mortality using Swedish life tables for the period 1901–70

It is seen in Fig. 8 that there is a strong linear relationship between A and $\hat{q}_1$ and that A at all levels is nearly the same as $\hat{q}_1$. The linear relationship is virtually the same for the two sexes. From an empirical point of view, this also means that A can be considered a linear function of infant mortality $\hat{q}_0$. Consequently, one can interpret A as an index of the level of childhood mortality.

The role of B is to position $q(0,\mathbf{e})$ appropriately above $q(1,\mathbf{e})$. More specifically, if B=0 then for any positive A and C, $q(0,\mathbf{e}) = 0.5$. Hence, B is an index of where $q(0,\mathbf{e})$ is positioned in the interval between $q(1,\mathbf{e})$ and 0.50. (Only in exceptional cases would infant mortality exceed $q_0 = 0.5$.) At ages above one year, B is of no significance to $q(x,\mathbf{e})$. Based on this, it would seem that B should be highly correlated with q_0-q_1. We see in Table 5, however, that B has a higher correlation with the ratio $\hat{q}_1/\hat{q}_0$ than with $\hat{q}_0-\hat{q}_1$. One could interpret B as an infant mortality shape parameter.

C is decisive for the level as well as for the shape of the model age pattern of mortality. It will be seen (Fig. 10) that C, as a function of $\hat{q}_0$, behaves in a fairly linear manner that is somewhat different for the two sexes. As such it can be interpreted more as a level than as a shape parameter of childhood mortality.

In passing, one will notice in Table 4 that the estimates of A and C change monotonically with the level of infant mortality. Because B also is decisive for the shape of the curve below age 1, it does not always decrease with decreasing levels of infant mortality. Nevertheless, from a general point of view, estimates of A, B and C decline with declining level of childhood mortality.

5.5. *The numerical behavior of the adult parameters*

It will be noted (Table 2) that the parameter estimates for the complete model (1.12) behave just as nicely as the ones for the childhood component (5.1). As the level of mortality decreases, estimates of the parameters A and C decrease monotonically (albeit with slightly different values than shown in Table 4). The estimates of B decline with declining levels of q_0 except for males for the period 1951–1960.

With respect to estimates of the remaining parameters D, E, F, G, and H, it is clear that even though they do not always change monotonically with changing levels of mortality, they follow a numerical pattern which agrees very well with their interpretations.

For both males and females, the estimates of D behave similarly, i.e., they change smoothly as the level of middle life mortality changes. (The increase in D relative to the period 1901–1910 can most likely be ascribed to the influenza pandemic that took place about 1915.)

The dispersion of the accident hump, as measured by $E' = 1/E$, increases as the size of the hump decreases. This is also what we would expect; for the more pronounced the accident hump, the smaller is its dispersion.

With respect to the location of the hump, as measured by F, it falls relatively slowly for males. For females the location of the hump has changed quite a bit over the decades and, as in the case of males, it falls over time.

The level of adult and late adult mortality, as measured by G, decreases monotonically with the level of mortality for males but displays no such trend for females.

The Gompertz parameter H is almost constant ($H = 1.1$) for all life tables but actually increases steadily with increasing life expectancy for males. (In the literature, the Gompertz parameter is sometimes referred to as demographically invariant.)

It will be seen that there is somewhat more regularity in the numerical behavior of the estimated parameters for males than for females. This is also reflected by the correlations between the parameters (Table 3). Notice that, on the whole, the parameter correlations are smaller for females than for males.

This finding is confirmed by earlier studies (Hartmann (1981, pp. 59–82)). When studying how middle life mortality was related to childhood and adult mortality in a large number of empirical life tables, it was found that, especially for females, there was a very weak relationship between middle life mortality (ages 15 to 35) and mortality outside this age range. A similar weak relationship between middle life mortality and mortality outside this age range (especially for females) can be observed for the Swedish life tables. The correlations in Table 3 show that the parameters $A, \ldots, H$ reflect this common feature of mortality. It is perhaps fitting to note that the magnitudes of the correlations (Table 3) also reflect the confidence one should attach to model life tables as tools for indirect estimation of mortality.

5.6. The potential uses of Heligman and Pollard's model

The well behaved nature of the parameter estimates (Table 2 and Table 4), paired with the fact that (1.12) gives unusually good fits to time-series of Australian and Swedish life tables suggests that it is an ideal and useful model for making population projections and for making demographic simulations where there is a need for a model of mortality at all ages. Although actual construction of model mortality curves is outside the scope of the present paper, it is clear that (1.12) is a flexible and highly useful continuous parametric model that could be used for modeling mortality curves. Model curves generated by (1.12) would have several obvious advantages. They are (1) easy to input, (2) smooth, (3) accommodate for changing age patterns of mortality as the level of mortality changes, (4) allow for investigating how improvements in given parts of the mortality curve would affect the age distribution of survivors, and the remaining life expectancy at given ages, and, in my view, (5) can be tailored to represent any given timeseries of national or regional life tables.

There is no existing life table system that has these properties. Typically, existing model life tables (United Nations (1955 and 1982), Brass (1971), Carrier and Hobcraft (1971), Ledermann (1969), Coale and Demeny (1966), Zaba (1979), Ewbank et al. (1983)) are (1) discrete representations of mortality, (2) time consuming to work with, (3) intractable to modify so as to allow for experiments where parts of the model survival tabulations are changed, and (4) they cannot be selected so that they are truly representative of a national age pattern of mortality (they are averages of several national age patterns of mortality). The Brass logit life table system, and its four-parameter modifications (although much more flexible than traditional model life tables), by and large, share these inadequacies.

6. Conclusions

Attempts to model mortality curves can be traced back to about 1870 when the Danish statistician and actuary Thiele proposed a law of mortality for all ages. It is based on the assumption that deaths naturally fall into three distinct categories, namely those of childhood, middle life, and adult life. Wittstein, a German statistician and actuary, proposed a similar law about 1883. It was his assumption that deaths could be divided into just two categories; childhood and adult ages. Wittstein's law, however, does not give fits to mortality curves that are competitive with those obtained from Thiele's. Over the years, the pioneering works of Thiele and Wittstein have received little more than marginal attention in demographic circles.

Recently, however, Heligman and Pollard have proposed a new law of mortality that reflects the ideas of Thiele. This model gives unusually close fits to empirical mortality curves. The parameters, which can be demographically interpreted, change relatively smoothly from one age pattern of mortality to another. For this reason it is an ideal demographic model for generating model curves for use with population projections and in other work where there is a need for a continuous model of mortality at all ages.

7. References

Brass, W. (1971): On the Scale of Mortality in Biological Aspects of Demography. London.

Carrier, N. and Hobcraft, J. (1971): Demographic Estimation for Developing Nations. London.

Coale, A.J. and Demeny, J. (1966): Regional Model Life Tables and Stable Populations. Princeton.

Cramer, H. (1963): Mathematical Methods of Statistics. Princeton, p. 354.

Elston, J.S. (1923): Survey of Mathematical Formulas that have been used to Express a Law of Mortality. The Record of the American Institute of Actuaries, Vol. 12, pp. 66–86.

Ewbank, D.C., Gomez de Leon, J.C., and Stoto, M.A. (1983): A Reducible Four-parameter System of Model Life Tables. Population Studies, Vol. 37 (1), pp. 105–127.

Hartmann, M. (1981): Laws of Mortality on the Logit Scale. Department of Statistics, University of Lund (Working paper).

Hartmann, M. (1983): Past and Recent Attempts to Model Mortality at all Ages. Department of Statistics, University of Stockholm (Working paper).

Heligman, L. and Pollard, J.H. (1980): The Age Pattern of Mortality. Journal of the Institute of Actuaries, Vol. 107, Part I, pp. 49–80.

Hendersen, R. (1915): Mortality Laws and Statistics. London.

Hoem, J.M. (1983): The Reticent Trio: Some Little-known Discoveries in Life Insurance Mathematics by L.H.F. Oppermann, T.N. Thiele and J.P. Gram. International Statistical Review, 51.

Hooker, P.F. and Longley-Cook, L.H. (1953): Life and Other Contingencies. Cambridge.

Jörgensen, N.R. (1913): Grundzüge einer Theorie der Lebensversicherung. Jena.

Ledermann, S. (1969): Nouvelles tables-types de mortality. Travaux et Documents, Cahiers, No. 53. INED, Paris.

Smith, D. and Keyfitz, N. (1977): Mathematical Demography. Springer-Verlag, New York.

Thiele, T.N. (1871): En Mathematisk Formel for Dödeligheden prövet paa en af Livsforsikringsanstalten af 1871 benyttet Erfaringsräkke. Copenhagen.

Thiele, T.N. (1872): On a Mathematical Formula to Express the Rate of Mortality throughout Life. Journal of the Institute of Actuaries, XVI, pp. 313–329.

Thiele, T.N. (1903): Theory of Observations. London.

United Nations (1955): Age and Sex Patterns of Mortality. Model Life Tables for Underdeveloped Countries.

United Nations (1982). Model Life Tables for Developing Countries. Population Studies, 77. New York.

Wittstein, J. (1883): The Mathematical Law of Mortality. Journal of the Institute of Actuaries, XXIV, pp. 153–173.

Wolfenden, H.H. (1942): The Fundamental Principles of Mathematical Statistics. Actuarial Society of America.

Zaba, B. (1979): The Four-parameter Logit Life-table System. Population Studies, 33 (1), pp. 79–100.

Received April 1986
Revised January 1987

Journal of Official Statistics
Vol. 3, No. 1, 1987, pp. 37–43

The Earliest Statistical Tables in China

Mu Zhu[1]

Abstract: This paper discusses the embryonic forms, the compilations, and the evolution of statistical tables from Shang Dynasty (1 600–1 700 B.C.) to West Han Dynasty (206–8 B.C.) in China. This paper emphasizes the work of the Chinese historian Sima Qian who compiled ten chronological tables in 91 B.C., the earliest statistical tables ever known in China. These tables consist of a combination of words and figures and provide basic information about the Dukedoms under the reign of Han Dynasty emperors. Sima Qian laid down explicitly the basic forms of statistical tables including headings, item titles, units of measurement, sequence of events, shape of table, and the name of the indicator. There seems to be little difference between these tables and the modern ones as far as the form is concerned. Sima Qian also put forward some theories about statistical tables. The theories and techniques of statistical tabulation created by Sima Qian have had a great influence on the development of science and culture in China.

Key words: Statistical tables, statistical data; *Book of History; Records of History;* chronological table, Sima Qian; tortoise shell inscription; tax rating.

Statistical tables are an important element in the presentation and analysis of statistical data. These tables are widely used for collecting, processing, analyzing, storaging and transmitting data. Statistical tables make it possible to use large amounts of figures in scientific research and economic management.

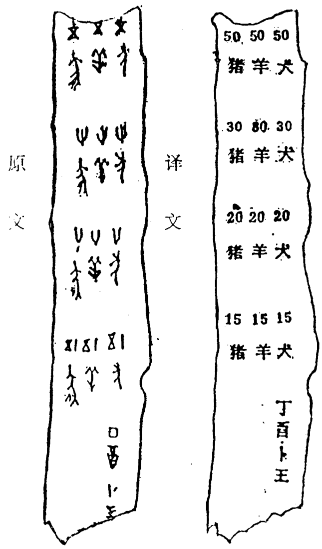
In this paper I will try to give some facts concerning the embryonic forms and the evolution of statistical tables in China.

1. Embryonic Forms of Statistical Tables in China

In China, rudiments of statistical tabulation can be traced back to the Shang Dynasty (16th – 17th century B.C.). Among the tortoise shell inscriptions of that period unearthed in 1899 near Anyang, Henan Province, there was a record of sacrificial livestock used by the king as offerings.[2]

[1] Chief, Teaching and Research Section, First Branch College of the Chinese People's University, Beijing, China.

[2] Compilation of the First Part of the Engraved Writings Unearthed in the Yin Xuins, Vol. 3, Chap. 6, p. 23.



The original inscription, its transcription into modern Chinese characters and its English translation are provided below.

50 pigs	50 sheep	50 dogs
30 pigs	30 sheep	30 dogs
20 pigs	20 sheep	20 dogs
15 pigs	15 sheep	15 dogs

Divination of the Shang King,
Year of Ding-You

It can be seen that, in this record, attention was paid to the classified arrangement of the data.

During the Warring States Period (475–221 B.C.), there were two texts referring to the use of statistical tables, namely, *Zhifangshi* and *Yu Gong*. The former was a chapter of *Rites of the Zhou Dynasty*, where it was written that "the *zhifangshi* was an official in charge of charts and household registration and must be well informed on the country's figures about land, mountains, rivers, cities and towns, population, finances, grain, and livestock; he must let everybody know where the advantages and disadvantages were, so that all the nine-state country might strive for common benefit."[3] The charts referred to here were primitive statistical tables.

In *Zhifangshi*, some general information was provided about the nine states in ancient China: their mountains, cities, rivers, lakes, products, population, livestock and crops. The eight categories mentioned above were displayed one by one in the sequence of the nine states. Of special interest is the sex ratio of the population in the nine states. The following figures are provided:

State	Sex ratio (male: female)	State	Sex ratio (male: female)
Yang	2:5	Yong	3:2
Jing	1:2	You	1:3
Yu	2:3	Ji	5:3
Qing	2:2	Bin	2:3
Yan	2:3		

Yu Gong[4] is an article in the *Book of History*, which is said to have been written by a Qin writer in the Warring States Period. It describes how matters stood in China in the period when Yu (the sovereign of the country who lived about the 21st century B.C.) fought the floods. The text also divides the country into nine states during the Xia Dynasty. The descriptions are similar to those in *Zhifangshi*. What is unique about this book is its classification of the farm land according to its fertility, as well as the taxes imposed on different states. Nine classes of land and taxes are identified.

Though statistical tables are not used in the below two texts, their expositions and their method of putting data together represent un-

[3] See note 62 of "Zhifangshi", Rites of the Zhiu Dynasty, Vol. 8.

[4] "The Taxes During Emperor Yu's Reign", in the Book of History.

State	Land fertility rating	Tax rating
Ji	upper upper	upper upper and middle upper
Yan	lower middle	lower middle
Qing	upper upper	upper middle
Xu	middle upper	middle middle
Yang	upper lower	middle lower
Jing	middle lower	lower upper
Yu	upper middle	upper and middle upper
Liang	upper lower	upper lower, middle lower, and lower lower
Yong	upper upper	lower middle

doubtedly the first step toward the development of tables later on.

2. The First Statistical Tables Compiled in China by Sima Qian in 91 B.C.

In his *Records of History,* completed in 91 B.C., Sima Qian (145–90 B.C.) of the Han Dynasty provided ten statistical tables: a genealogical table of the Three Ancient Dynasties (Xia, Shang and Zhou), a chronological table of 12 dukes and princes, a chronological table of six states (dukedoms), a month-by-month table of events between the Qin and Chu States, a chronological table of dukes since the founding of the Han Dynasty, a chronological table of loyal ministers having rendered distinguished service, one of dukes during the period of Emperor Gaozu of the Han Dynasty, a chronological table of dukes of the Xiaohui-Xiaojing period, a chronological table of dukes since the Jianyuan period (one of the periods of reign of Emperor Wu of the Han Dynasty), a chronological table of princes of the same period, and a chronological table of noted generals and ministers since the founding of the Han Dynasty[5]. These are the earliest tables of statistics known in China and preserved up to our days. In the Appendix on page 43 we present one of the chronological tables in its original form and Table 1 on the following page shows it in English version.

Let us have a look at the state of Bian. The Prince of Changsha had rendered distinguished service to the Dynasty. His son Wu Qian, therefore, was granted the title of Duke of Bian and enfeoffed with 2 000 households. He ruled the dukedom for 37 years (194–157 B.C.), namely, seven years during the reign of Emperor Xiaohui, the eight years of that of Emperor Gao, and 22 years during that of Emperor Xiaowen). He was succeeded by his son Wu Xin, who ruled the dukedom for six years and was succeeded in his turn by Wu Guangzhi, who was duke for 39 years. Wu Qianqiu, the great grandson of Wu Qian, ruled the dukedom for less than one year. His state was abolished because he had no son. The dukedom existed for a total of 82 years.

If we list only the number of households given to the dukes and the number of years their dukedoms existed, we will get the following table:

State*	Number of households given by the emperor	Number of years the state existed
Bian	2 000	82
Dai	700	83
Pingdu	1 000	48
Wu	500	69
Zhongyi	600	43
Leping	600	49
Chengtao	500	19
Liling	600	8
Yangxin	2 000	28
Zhi	10 000	40

* Here only ten of the 93 dukedoms are taken for example.

[5] Sima Qian, Records of History, China Publishing House, Vol. 19, p. 978.

Table 1. Chronological Table of Dukes of the Xiaohui-Xiaojing Period

State	Bian	Dai	Pingdu
Merits for enfeoffment	Prince of Changsha, enfeoffed with 2 000 households	Chancellor of Changsha, enfeoffed with 700 households	Former general of the Qi State, having passed over to the Han Dynasty and being enfeoffed with 1 000 households
Reign of Emperor Xiaohui, 7 years (194–188 B.C.)	Sept., the 1st year, Wu Qian was granted the title of duke 7*	April, the 2nd year, Lichang was granted the title of duke 6	June, the 5th year, Liu Dao was granted the title of duke 2
Reign of Emperor Gao, 8 years (187–180 B.C.)	8	2 The 3rd year, Xi was granted the title of duke	8
Reign of Emperor Xiaowen, 23 years (179–157 B.C.)	22 The last 7th year, Xing was granted the title of duke 1	15 The 16th year, Pengzu was granted the title of duke 8	2 The 3rd year, Cheng was granted the title of duke 22
Reign of Emperor Xiaojing, 16 years (156–141 B.C.)	5 The 6th year, Guangzhi was granted the title of duke 11	16	14 The last 2nd year, duke Cheng committed crimes and his dukedom was abolished
From the Jianyuan period to the 6th year of the Yuanfeng period, 36 years (140–105 B.C.)	29 The 5th year of the Yuanding period, Qianqiu had not paid duly the sacrificial costs. The dukedom was abolished because he had no son.	30 The 1st year of the Yuanfeng period, duke Zhi, governor of Donghai, was guilty of mobilizing troops without authorization, and had his head cut off. His dukedom was abolished.	
After the Taichu period (104–B.C. –)			

* The figures listed in the upper and lower right corner of each square show the length of the reign of the respective dukes during the period.

The Chronological Table of Dukes of the Xiaohui-Xiaojing Period is the seventh of the ten tables compiled by Sima Qian. It describes the basic events and conditions of the 93 dukedoms created during the period between the first year of Emperor Xiaohui's reign and the 6th year of the Yuanfeng period of Emperor Wu's reign. Here only three of the 93 dukedoms are given as examples.

The tables made by Sima Qian already had the basic features of modern statistics. Sima Qian laid down explicitly the basic forms of statistical tables including headings, item titles, units of measurement, sequence of events, shape of the tables and the name of the indicators, etc. The tables are enclosed with thick lines on the margin, and thin lines are used inside. Top and bottom lines combined with horizontal and vertical ones inside make up a rectangular form. The old Chinese writing system follows the rule "from the top to the bottom, from the right to the left," and this is also the way the words and figures are written in the tables. The heading is written to the right side of the table, and the item titles of the vertical and horizontal columns are written between the first two horizontal lines and the first two vertical lines, respectively. The objects being observed are classified according to their different nature, and they are listed under different titles, such as name of the state, number of households given to the duke, number of years of the duke's reign, and the years of establishment and abolition of the state. Following this format, and following a certain sequence, the basic conditions and events of each state are recorded in the tables. In some of these tables, the names of the states are given in vertical columns, each state occupying one column. The number of households and the rise and fall of the states are also recorded. The time of the events is given in the horizontal columns, each period occupying a column. After having looked at the tables, you have an idea of the changes that took place in all these states.

3. The Theory About Statistical Tables Put Forward by Sima Qian

Sima Qian also put forward his theory about statistical tables. He said: "Confucians used to quote out of contexts, and persuasive talkers often indulged in exaggeration. They did not seek to study the whole story comprehensively. Almanac writers only studied the years and the months, astrologers stressed predestination and fate, and heraldry specialists only made records of posthumous titles of emperors of different generations. All these records were simple and one-sided, making it difficult to view various important historical facts in a comprehensive way. For this reason, I have worked out a chronological table of 12 dukes starting from the Gonghe period of the Zhou Dynasty up to the Confucian period. I drew up tables of general information concerning the rise and fall of the states, events which had been used as allegory by scholars in the *Spring and Autumn Annals* and *Conversations from the States*. In this work, I avoided superfluous wording and tried to present the essentials, and I offered it as reference for those who engaged in academic research and political administration."[6]

"I therefore followed the historical record of the Qin Dynasty and the example of the *Spring and Autumn Annals*. Starting from King Yuan of the Zhou Dynasty, I recorded all the events that happened in the six states up to the reign of the Second Emperor of the Qin Dynasty, covering a period of 270 years in total. The causes for the rise and fall of the states are explained. This record is left to future scholars for reference."[7]

Here Sima Qian pointed out that statistical tables are useful for:

1) avoiding superfluous wording and depicting what is essential in a concise way;
2) organizing systematically and arranging rationally different important data, in order to make it easier to observe and compare them, and

[6] Sima Qian, Records of History, China Publishing House, Vol. 14, p. 511.
[7] Sima Qian, Records of History, China Publishing House, Vol. 19, p. 978.

3) facilitating the comprehensive study of the whole process of the march of events from the beginning to the end, as well as the laws governing its development.

Sima Qian's purpose in drawing up his tables was to provide general information about the rise and fall of the states as well as the management of state affairs. Such a theoretical view is still of importance today.

The theory and techniques of statistical tabulation created by Sima Qian have had a great influence on the development of science and arts in China, and have enabled a good many of Chinas famous scholars and practical workers to make important contributions in their academic and scientific research and state administrative activities.

4. Some Evaluation Concerning the Statistical Tables Drawn Up by Sima Qian

Scholars of the past spoke highly of the statistical tables drawn up by Sima Qian. Liu Zhiji, a historian of the Tang Dynasty, said: "Looking at the tables drawn up by Sima Qian, I can see what happened throughout thousands of miles and nine generations of sovereigns, all condensed in an orderly way in a small table of a few square inches. Have a look at the tables, and you will know exactly what you want to know."[8]

Zheng Qiao, another noted historian of the Song Dynasty, said: "In drawing up these tables, Sima Qian made use of only six of the 60 combined pairs of the ten Heavenly Stems and the twelve Earthly Branches, beginning from the first of the ten Heavenly Stems, to record events chronologically. The other 54 pairs were not used. With only a few words he managed to convey what was essential in the events, which means that once the essential thing is grasped, everything falls into place."[9] Then he talked about the importance of statistical tables. He said: "You can do without statistical tables if you just want to indulge in empty talk. But you cannot do without them if you really want to accomplish anything." And he ascribed the failure of many scholars and politicians after the Han Dynasty to their inability to use statistical tables.

5. Summary

In his *Records of History*, completed in 91 B.C., Sima Qian (145–90 B.C.) of the Han Dynasty provided ten statistical tables. Examples are: a genealogical table of the Three Ancient Dynasties, a chronological table of six states, etc. These are the earliest tables of statistics ever known in China. Sima Qian also put forward his theory about statistical tables in his *Records of History*. He pointed out that statistical tables serve as an important means for the presentation and analysis of statistical data, facilitating the study of the whole process of the march of events from the beginning to the end, as well as the laws governing its development.

Received February 1986
Revised October 1986

[8] Liu Zhiji, Notes and Reflections on a Variety of Matters.
[9] Zheng Qiao's illustrative plates in Comprehensive History.

Appendix

A Chronological Table of Dukes in the Huijing Period in the Original Chinese

惠景间侯者年表第七

国名	侯功	孝惠七	高后八	孝文二十三	孝景十六	建元至元封六年三十六	太初已后
便 【索隐】汉志县名，属桂阳，音鞭。	长沙王子，侯，二千户。	七 元年九月，顷侯吴浅元年。	八	二十二 一 后七年，恭侯信元年。	五 十一 前六年，侯广志元年。	二十八 元鼎五年，侯千秋坐酎金国除。	
轪 【集解】音大。【索隐】轪音大，县名，在江夏也。	长沙相，侯，七百户。	六 二年四月庚子，侯利仓元年。【索隐】汉书作（轪侯朱仓）故长沙相。	二 六 三年，侯豨元年。	十五 八 十六年，侯彭祖元年。	十六	三十 元封元年，侯秩为东海太守，行过不请，擅发卒兵为卫，当斩，会赦，国除。	
平都 【索隐】县名，属东海。	以齐将高祖三年降，定齐，侯，千户。	三 五年六月乙亥，孝侯刘到元年。【索隐】故齐将，已上孝惠时三人也。	八	二 二十一 三年，侯成元年。	十四 后二年，侯成有罪，国除。		