Anthrop. Közl. *27;* 129—137. 1983.

DISTRIBUTION OF THE PARAMETERS OF THE GOMPERTZ AND WEIBULL FUNCTIONS FITTED TO THE DEATH RATES

by J. Izsák

Department of Biology, Teachers' College, Szombathely

Abstract: We have fitted $a_i \cdot e^{b_i \cdot t}$ shaped Gompertz function and $a_i \cdot t^{b_i}$ shaped Weibull function to the $q_i^{(i)}$ death rate of 100 different death causes based on the vital statistics of the United States. According to the $d(= r^2)$ determination coefficient, in both cases the fitting could be considered as good. The total of the 100 b_i parameters in case of the Gompertz function according to the χ^2 test at a 5% level could be considered as of normal distribution, in case of the Weibull function it is not. The histogram of the logarithms of the a_i parameters is more irregular, but also unimodal. The distribution of definite mode implies that some sort of statistical principle is valid in the formation of the a_i and b_i stochastic parameters. The close to normal or close to lognormal distribution of the b_i parameter can be related to our earlier observation that the death cause frequencies follow a lognormal, or at least close to lognormal distribution.

Key words: Gompertz function, Weibull function, normality, lognormality, distribution of parameters.

Introduction

The incredible fast rate of growth of the average lifespan of the hominids and then of mankind is partially due to physiological, and partially — in historical times practically only — to social factors. Thus the cephalization of an exceptional rate and size, then the formation of a social environment lifted mankind to among the longest-living species (Cutler 1976). It is clear that the average life expectancy of 70—80 years of today stands closer from the human biological point of view to the natural than the earlier one cut about half by the bad living conditions and epidemics. As the vital statistics show, the first biological signs of ageing are manifested exactly in the period (age 30—35) which earlier was close to the limit of the average human lifetime. In this period, from the age of 35, age becomes a dominant factor in death dynamics (Gavrilov et al. 1978).

More concretely, there are a lot of data supporting the idea that in the often accepted form $A + a \cdot e^{bt}$ of the $\lambda(t)$ hazard function (Gompertz—Makeham model), the A term embodying the background hazard, as compared to the exponential term, will be negligible from the age of 30—35 (Gavrilov et al. 1978). From this age on, the death hazard could be modelled well by the $\lambda(t) = a \cdot e^{bt}$ exponential function (Gompertz model).

The useability of the Gompertz function has a large demographic and gerontologic literature. It is important to follow it in certain areas of the anthropological investigations too (Acsádi and Nemeskéri 1970). There are less publications about the changing of the risk of different death causes due to age (KOHN 1963, LINZBACH 1975, UPTON 1977). And the ensemble of the $q_t^{(i)}$ death rates of the different death causes are hardly ever discussed as subject matters of investigations. The reason for this could partially be that these types of investigations are relatively far from the center of attention of classical demography, gerontology and epidemiology. The global review of the structure of death causes according to age — which is the subject of our publication too — will lead to paying more attention to the human biological points of view.

According to our earlier research, the change of certain parameters of the ensemble of the death causes due to age and the changes according to race and sex deserve attention from the human biological point of view. Because we found that the concentration of death causes calculated using the death cause categories of a great number shows changes characteristic of age (IZSÁK and JUHÁSZ-NAGY 1981, IZSÁK 1982), and its value is generally higher in case of men than in case of women (IZSÁK and JUHÁSZ-NAGY 1980, 1981). The difference is more pronounced in the white race than in the black race (IZSÁK and JUHÁSZ-NAGY 1981). (We have to note though, that the latter difference in the studied statistics could be due to the difference in way of life rather than to that in race.)

The question has arisen: does the total of death hazards have some sort of structure in the statistical sense (IZSÁK and JUHÁSZ-NAGY 1982a, 1982b)? We analyze the following two questions in our publication related to the above:

— What kind of formal characterization could be given for the set of graphs of the $q_t^{(i)}$ death rates between the ages of 35—79, that is, within the generally accepted domain of the Gompertz function. We wanted to study that to what extent are the Gompertz function of the form $\lambda(t) = a \cdot e^{bt}$ and the Weibull function of the form $\lambda(t) = a \cdot t^b$ suitable for the approximation of the concrete $q_t^{(i)}$ death rates. [Recently the use of Weibull's hazard function has become wide-spread too (GROSS and CLARK 1975).]

— Denoting the function fitted to the graph of the i death rate by $a_i \cdot e^{b_i t}$ and by $a_i \cdot t^{b_i}$, what is the distribution of b_i and of a_i?

This latter question could be related to our earlier observation that normal distribution could be fitted well to the logarithms of the frequencies of death causes (IZSÁK and JUHÁSZ-NAGY 1982b). It seemed to be important the statistical study of the set of death rates also because without this it would be difficult to interpret the secular changes of the concentration of death causes. If we talk about the redistribution of the suppressed frequencies of death causes, or about the changes concerning equally the elements of the ensemble of death causes, the simultaneous consideration of the considered death causes of a great number (sometimes several hundreds) is possible only on the basis of statistical experiences and on the basis of such models.

Material and methods

We carried out our studies on the 1974 and 1975 death cause statistics of the United States (Vital Statistics of the U.S. see at the literature), on subpopulation of white males. This material had often been used in our earlier works. Fitting was done at the age interval of 35—79, using those of the earlier considered 161 death cause categories of the ICD (IZSÁK and JUHÁSZ-

NAGY 1981) in which the number of death cases within the age group of 35—39 for the years 1974 and 1975 was at least 5. Thus the number of ICD categories included in the study was exactly 100. We calculated the $q_t^{(i)}$ death rates for the 35—39, 40—44, etc. age groups so that the number of death cases due to i death cause within the *five-year* age group was divided by the number of the midyear population of the middle year given in thousands. The actuarial estimates of the $q_t^{(i)}$ death rates calculated (GEHAN 1969) was multiplied by 100 for graph plotting reasons, then, after logarithmization we fitted straight line by the method of least squares in case of the Gompertz function to the pairs of points $(t; \ln q_t^{(i)})$ $(t = 37.5; 42.5; \ldots)$, in case of Weibull function to the pairs of points $(\ln t; \ln q_t^{(i)})$. The slope of the fitted straight line can be considered as the exponential parameter of the fitted curve in both cases. Also in both cases the axial intercept of the straight line is $\ln a_i$, which is actually $\ln 5 \times 10^5 = 13.12$ times as big as the real $\ln a_i$ (see above). In our conclusion this has no importance.

We judged the goodness of fitting by the determination coefficient, that is, by the square of the linear correlation coefficient (EZEKIEL and FOX 1959). We considered this simple index suitable for our investigations of informative and general character.

Results and discussion

The graphs according to age of the logarithmized $q_t^{(i)}$ values — as expected — showed curves close to the straight line or perhaps a bit concave curves. On Figure 1. the semilogarithmic plotting of the death rate graphs of the neoplasms can be seen. (Due to this transformation the shape of the exponential curve will be a straight line.) Others have also observed similar sets of curves, even if not in such a fine division of death causes. According to the d determination coefficient calculated after the fitting of the Gompertz and Weibull functions, the fitting of both functions could be said good. As regards the 100 ICD categories of the study the average of the d coefficient in case of the Gompertz function was 0.888, in case of the Weibull function it was 0.971. The fitting of the former function seemed better than the other in 45 cases, and the opposite — the better fitting of the Weibull function — was observed in 55 cases. Thus the measure of fitting cannot support clearly one or the other type of curve. Perhaps the Gompertz function can be supported by the fact that fitting of the Weibull function is better generally when the $q_t^{(i)}$ graph is concave, which usually is the result of civilizational effects. It is necessary to analyse further though that with what constellation of the parameters could the (total) Gompertz curve exist as the sum of the graphs of near exponential growth (see LOHMANN 1978).

The histograms of the b_i parameters in the exponents of the fitted exponential and power functions can be seen on Figures 2a) and b). It can be stated that the b_i parameters follow a distribution at least near to the normal in both cases. Some irregularity can be seen only among the greater b_i parameters of the Weibull function. The χ^2 value calculated for the normality study in case of the Gompertz function was 10.0 after the unification of the two lower and the three upper frequency classes, and in case of the Weibull function it was 14.8 after the unification of the three lower and the two upper frequency

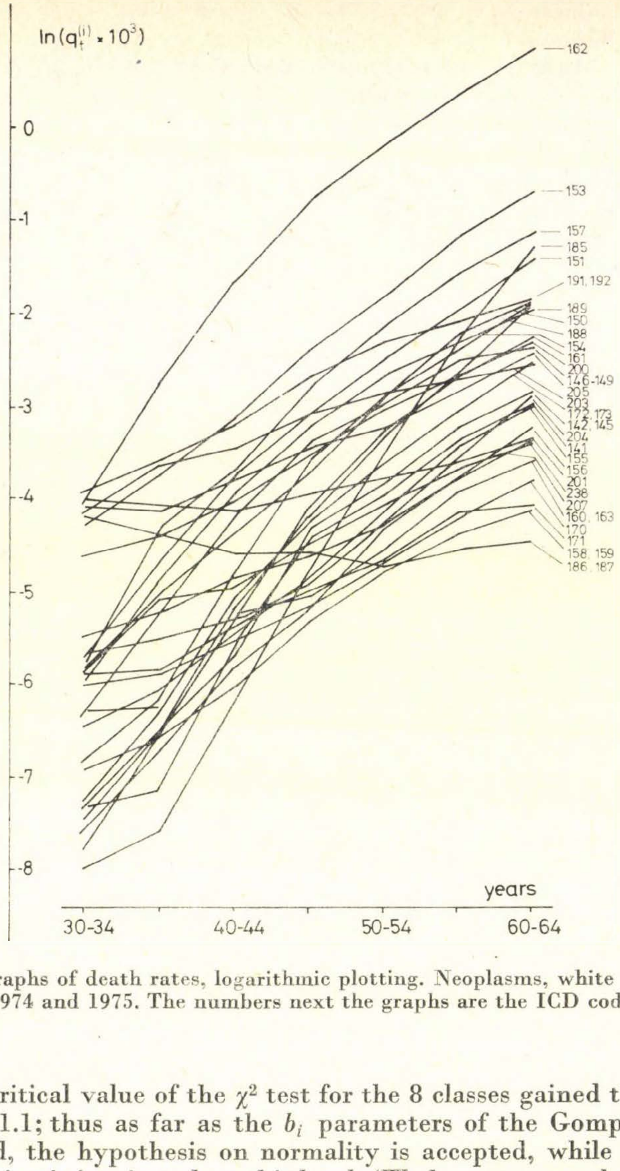

Fig. 1. Set of graphs of death rates, logarithmic plotting. Neoplasms, white males in the US, 1974 and 1975. The numbers next the graphs are the ICD codes

classes. The critical value of the χ^2 test for the 8 classes gained this way at the 5% level is 11.1; thus as far as the b_i parameters of the Gompertz functions are concerned, the hypothesis on normality is accepted, while in case of the Weibull function it is rejected on this level. (We have to note that in the latter case the essentially greater χ^2 value could be mostly due to the especially small frequency of the [7—8] interval. Therefore the normality of the b_i parameters of the Weibull functions should not be rejected totally!)

The histograms of the logarithms of the a_i parameters (the axial intercepts of the straight lines) can be seen on Figures 3a) and b). As we have already mentioned, only the categories with a frequency of at least 5 within the age group of 35—39 were included in the fitting study. This is the explanation for the fact that the Gompertz histogram (the more regular one) is slightly askew

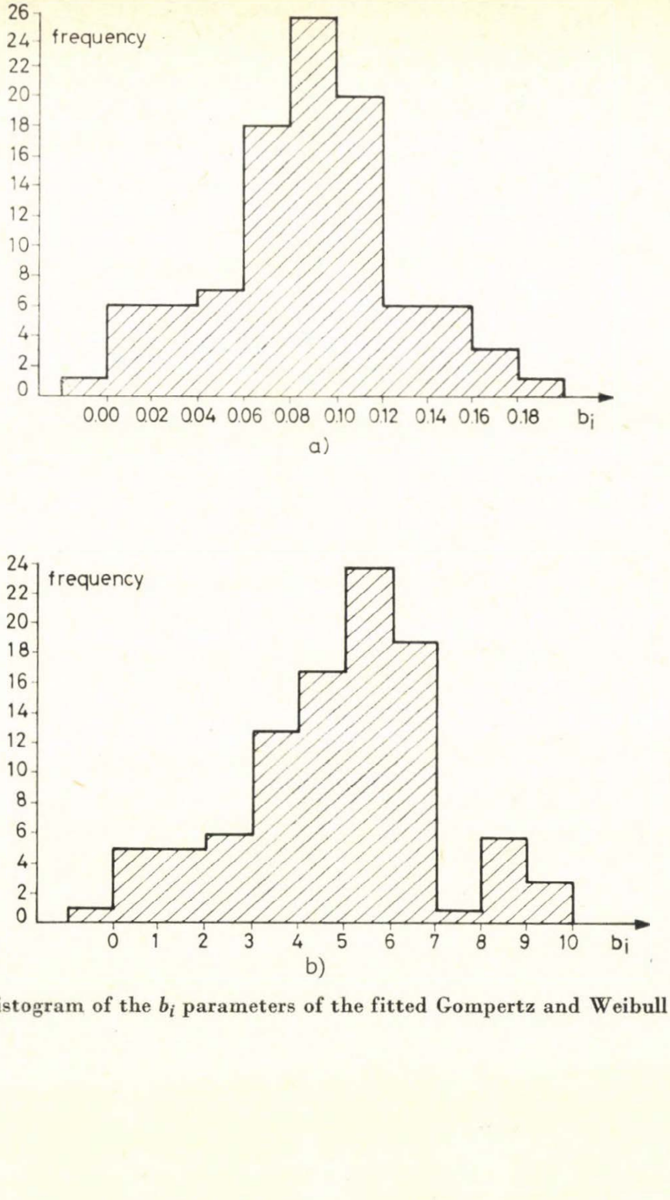

Fig. 2. **The histogram of the b_i parameters of the fitted Gompertz and Weibull functions**

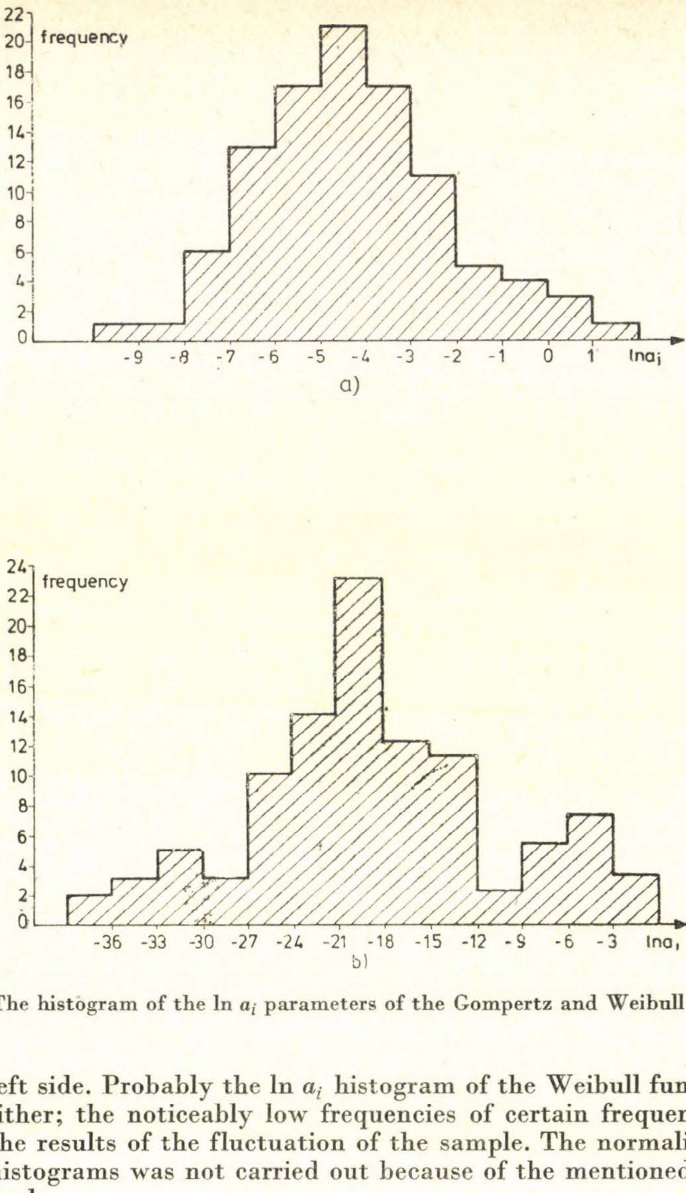

Fig. 3. The histogram of the ln a_i parameters of the Gompertz and Weibull functions

from the left side. Probably the ln a_i histogram of the Weibull function is not atypical either; the noticeably low frequencies of certain frequency sections could be the results of the fluctuation of the sample. The normality study of the ln a_i histograms was not carried out because of the mentioned truncation of the sample.

Evaluation of the results

As we have seen, both the Gompertz and the Weibull functions are suitable for the approximation of the death rates within the age interval of 35—39, and for the modelling of the set of the hazard functions $\lambda_i(t)$ for certain purposes.No matter which model we choose, the fitted curves move apart form each other at a faster rate than the linear. This means that the concentration of death causes — or the calculability of the death causes — will increase with age.

The normal, or close to normal distribution of the b_i parameters in case of the Gompertz model is undoubtable. This could not be said about the Weibull model. It is possible though that there are some disturbing factors behind the deviation from the normal distribution. Nevertheless, in both cases the normal or near to normal distribution implies that *we should look for a kind of common mechanisms behind the formation of the b_i slope parameters.* Without this, we would have to count perhaps with the even distribution of the b_i parameters, and by no means with an unimodal distribution. Practically the same thing could be said about the distribution of the a_i and the $\ln a_i$ parameters. The definite mode of the histograms implies a statistical regularity which forms the size of the parameters.

As far as the irregularities of distribution of the parameters of the fitted Weibull functions are concerned, they support — even if only a little bit — the Gompertz model.

The normal, or close to normal distribution of the b_i parameters is most probably related to the already described phenomenon that the lognormal distribution can be fitted well to the death cause frequencies (see above). This relationship could be based on the fact the logarithms of the $\lambda_i(t)$ hazard function values directly related to death cause frequencies are equal to the sums of the logarithms of the a_i parameters and of the $t \cdot b_i$ values or $\ln t \cdot b_i$ values.

In case of the Gompertz model:

$$\ln \lambda_i(t) = \ln a_i + t \cdot b_i,$$

in case of the Weibull model:

$$\ln \lambda_i(t) = \ln a_i + \ln t \cdot b_i.$$

Nevertheless, for the clarification of the question more investigation is needed.

Summary

The global investigation of death hazards is the task of epidemiology partially, and partially of human biology. The detailed epidemiological analysis of the group of more frequent illnesses and death causes, from the point of view of evolution studies could not lead to results, because they could hardly be taken into account when judging the basically biological questions due to the different effects of the social environment. The point of view which considers the widest possible circle of illnesses as the basis of investigations seems to be more promising. With this idea of investigation we have studied the statistical characteristics of the *total* of the death hazard functions.

We have established that the exponential increase due to age between the age of 35—39 of death rates, which is known in literature as Gompertz's law, is *divided* in the sense that exponential curves can be fitted well generally to the age-graphs of certain death rates. At the same time, the fitting of the

power function (Weibull model) had good results too. The slope parameter of the increase due to age of death hazards in case of the Gompertz functions could be considered as of normal distribution, in case of the Weibull function the normality hypothesis at the P = 5% level must be rejected. The normal, or close to normal distribution of the parameters supports the idea that there is a common background to the formation of the b_i slope parameters. Practically the same idea is raised when studying the distribution of the logarithms of the a_i parameters.

The normality of the b_i parameters and the lognormality of the a_i parameters most probably have a role in the phenomena described by us earlier, that lognormal distribution can be fitted well to the total of death cause frequencies.

As the continuation of our study, in the future we would like to investigate how the structure of the set of functions changes in case of different subpopulations.

HALÁLOKI RÁTÁKHOZ ILLESZTETT GOMPERTZ- ÉS WEIBULL-FÜGGVÉNYEK PARAMÉTEREINEK ELOSZLÁSA

Írta: IZSÁK JÁNOS

Összefoglalás

Az Amerikai Egyesült Államokbeli vitálstatisztika alapján 100 különböző halálok $q_t^{(i)}$ halálozási rátájához $a_ie^{b_it}$ alakú Gompertz-függvényt és $a_it^{b_i}$ alakú Weibull-függvényt illesztettünk. A d ($= r^2$) determinációs együttható szerint az illeszkedés általában mindkét függvény esetében jónak mondható. A 100 b_i paraméter együttese a Gompertz-függvény esetében χ^2 próba szerint 5%-os szinten normális eloszlásúnak tekinthető, a Weibull-függvény esetében viszont nem. Az a_i paraméterek logaritmusainak hisztogramja szabálytalanabb, de szintén unimodális. A határozott móduszú eloszlás arra utal, hogy az a_i ill. b_i kockázati paraméterek kialakulásában valamilyen statisztikai elv érvényesül. A b_i paraméterek normálishoz, ill. lognormálishoz közeli eloszlása kapcsolatba hozható azon korábbi megfigyelésünkkel, hogy a haláloki frekvenciák lognormális vagy legalábbis ahhoz közeli eloszlást követnek.

*

(A Magyar Biológiai Társaság Embertani Szakosztályának 1982. december 13-i szakülésén elhangzott előadás; közlésre beérkezett 1983. szeptember 8-án.)

LITERATURE

ACSÁDI, GY. and NEMESKÉRI, J. (1970): *History of Human Life Span and Mortality* Chap. I. — Akadémiai Kiadó, Budapest.

CUTLER, R. G. (1976): Nature of ageing and life maintenance processes. *In:* CUTLER, R. G. (Ed.) *Cellular Ageing: Concepts and Mechanisms.* Part I. — S. Karger, Basel.

EZEKIEL, M. and FOX, K. A. (1959): *Methods of Correlation and Regression Analysis.* Chap. 7. — Wiley, N. Y.

GAVRILOV, L. A., GAVRILOVA, N. S. and YAGUZHINSKI, L. S. (1978): The main regularities of animal aging and death viewed in terms of the reliability theory. — Zhurnal Obsh. Biol. *39;* 734—742.

GEHAN, E. A. (1969): Estimating survival functions from the life table. — J. Chron. Dis. *21;* 629—644.

GROSS, A. J. and CLARK, V. A. (1975): *Survival Distributions: Reliability Applications in the Biomedical Sciences.* Chap. 1. — Wiley, N. Y.

IZSÁK, J. (1982): Comparative study of diversity indices on mortality data. — Biológia *30;* 193—204. (in Hungarian, with an English Summary).

IZSÁK, J. and JUHÁSZ-NAGY, P. (1980): Investigation of diversity on Hungarian mortality statistics. — Ann. Univ. Sci. Budapestinensis, Sectio Biol. *22—23;* 35—45.
— (1981): Diversity studies on mortality data. Ann. Univ. Sci. Budapestinensis, Sectio Biol. (in press).
— (1982a): On the study of patterns of diseases. — Magyar Tudomány *27;* 37—43. (in Hungarian).
— (1982b): Studies of lognormality on mortality statistics. — Biometrical Journal *24;* 731—741.
KOHN, R. R. (1963): Human aging and disease. — J. Chron. Dis. *16;* 5—21.
LINZBACH, A. J. (1975): Altern und Krankheit. Ableitung einer neuen Alternstheorie auf der Grundlage der Polypathie. — Verhandlungen der Deutschen Gesellschaft für Pathologie *59;* 242—251.
LOHMANN, W. (1978): Eine aus dem Altersindex nach Ries folgende Modifikation der Gompertz-funktion und Approximierung der Absterbe-kurve. Zeitschrift für Alternsforschung *33;* 507—512.
UPTON, A. C. (1977) Pathobiology. *In:* FINCH, C. E. and HAYFLICK, L. (Eds): *Handbook of the Biology of Ageing.* — Van Nostrand Reinhold Company, N. Y.
Vital Statistics of the U. S. 1974, Vol. 2., *Mortality* (1978). U. S. Dep. of Health, Education and Welfare, Public Health Service, Govt. Print. Off., Washington, pp. 1—186—1—251.
Vital Statistics of the U. S. 1975, Vol. 2., *Mortality* (1979). U. S. Dep. of Health, Education and Welfare, Public Health Service, Hyattsville, pp. 1—186—1—251.

Author's address:
A szerző címe: DR. IZSÁK JÁNOS
Tanárképző Főiskola
Szombathely, Szabadság tér 4.
H-9701