RT&A # 01 (36)
(Vol.10) 2015, March

HUMAN LIFESPAN DISTRIBUTION WITH MAXIMUM LIFESPAN PARAMETER

Mark P. Kaminskiy
NASA Goddard Space Flight Center
•
8800 Greenbelt Rd, Greenbelt, MD 20771
mark.p.kaminskiy@nasa.gov , makaminsk@umd.edu

ABSTRACT

The problem of existence of maximum human lifespan is discussed. Assuming that the maximum human lifespan exists, a new lifespan distribution is suggested. In opposite to the popular lifespan distributions (Gompertz, Weibull, Extended Weibull, etc.) supported on the semi-infinite interval $[0, \infty)$, the suggested lifespan distribution is supported on the finite interval $[0, a)$, in which a is the non-random maximum lifespan. The suggested lifespan distribution was applied to three death rate datasets (Australia, France, and Switzerland) from the Human Mortality Database, for which the parameters of the suggested lifespan distribution were estimated. The fitted death rates have the high proportion of variance explained by the models ($R^2 \geq 0.96$), and the estimated maximum lifespan is about 200 years. A more adequate lifespan distribution might be a distribution having two competing risks – the risk of death from diseases, and the risk of death from "pure" aging.

INTRODUCTION

The existence of maximum human lifespan is of a great interest and debate to scientists as well as lay audience. This question also belongs to the wider problem of human ageing, which is investigated using multidisciplinary approaches, e.g., integrating population genetics methods with the principles of epidemiological and demographic investigation (Tan et al. 2004; Hayflick 2000).

An important methodological tool for ageing studies is the mortality data analysis (Tan et al. 2004). Note that any data analysis is based on the respective probabilistic model, which, in the case considered, is the lifespan distribution. Various distributions were suggested as the human lifespan models such as Gompertz, Weibull and Extended Weibull distributions (Tan et al. 2004; Weon and Je 2009).

It should be noted that all these distributions are supported on the semi-infinite interval $[0,\infty)$[1], which means that the lifespan (age at death) can take on any value from interval $[0,\infty)$. Using the distributions with infinite or semi-infinite support obviously makes it difficult to suggest a simple and non-random definition of the maximum lifespan. The common definition, which states that the maximum lifespan is "the maximum observed lifespan of a species" (Tan et al. 2004; Hayflick 2000), just reduces it to a realization of an unknown random variable. At most, it can be considered as a common sense point estimate of the maximum lifespan.

According to the suggested below definition, the maximum human lifespan is a non-random distribution characteristics, like the mean, median, mode, etc.

[1] In the case of normal distribution also used in the ageing studies, the interval is infinite, i.e. $(-\infty, +\infty)$ [4]

In the framework of lifespan distribution models approach, the question we pose is -- does the human lifespan, as a random variable, have a non-random limit? If it does, the respective lifespan distribution should have a finite positive support.

Below, we introduce a new lifespan distribution supported on the finite interval $[0, a)$, $a > 0$, in which a is the non-random maximum lifespan, and it is also one of the distribution parameters. Then, we consider some case studies demonstrating how the suggested distribution fits the mortality data for three countries having high life expectancy.

LIFESPAN DISTRIBUTION

The suggested distribution is based on the assumption that the human lifespan is a random variable having a nonrandom upper limit. The distribution is introduced through its death rate (mortality rate, hazard rate, of failure rate), which is given by

$$h(t) = \frac{b}{(a-t)^{\beta}}, \tag{1}$$

where t is the lifespan ($0 \le t < a$), parameter $a > 0$ is the maximum lifespan, and b and β are the other distribution parameters (one can call β the *shape parameter*).

Using the death rate (1), the main probabilistic functions related to our distributions can be obtained. The respective cumulative death rate is

$$\begin{aligned} H(t) &= b\int_0^t \frac{1}{(a-\tau)^{\beta}}\, d\tau \\ &= \frac{b}{(\beta-1)(a-t)^{\beta-1}} \end{aligned} \tag{2}$$

Using (2), the cumulative distribution function of lifespan can be written as

$$F(t) = 1 - exp\left[-\frac{b}{(\beta-1)(a-t)^{\beta-1}}\right] \quad , \tag{3}$$

and the survival function as

$$S(t) = exp\left[-\frac{b}{(\beta-1)(a-t)^{\beta-1}}\right] \tag{4}$$

The respective lifespan probability density function (pdf) is

$$f(t) = \frac{b}{(a-t)^{\beta}} exp\left[-\frac{b}{(\beta-1)(a-t)^{\beta-1}}\right] \tag{5}$$

Using Equations (4) and (5), the quantile of level p, or the 100pth percentile ($0 < p < 1$), can be written as

$$t_p = a - \left[\frac{1-\beta}{b}\ln(1-p)\right]^{\frac{1}{1-\beta}} \tag{6}$$

Using the pdf (5), one can find the respective mode as

$$t_{mode} = a - \left(\frac{\beta}{b}\right)^{\frac{1}{1-\beta}} \quad (7)$$

The mean, variance and other moments can't be found in closed forms. For example, the distribution mean can be written as

$$E(t) = \frac{a}{1-\beta}\left[\frac{b}{\beta-1}a^{\beta-1}\right]^{-\frac{1}{\beta-1}}\Gamma\left(\frac{1}{\beta-1}, \frac{\beta a^{\beta-1}}{\beta-1}\right) \quad (8)$$

or

$$E(t) = \frac{1}{1-\beta}\left[\frac{b}{\beta-1}\right]^{-\frac{1}{\beta-1}}\Gamma\left(\frac{1}{\beta-1}, \frac{\beta a^{\beta-1}}{\beta-1}\right)$$

where $\Gamma(a, b)$ is the incomplete gamma function, which is not easy to calculate. However, numerical evaluation of the distribution moments, based on their definitions is not difficult.

CASE STUDIES

The suggested lifespan distribution was applied to three datasets from the Human Mortality Database (HMD) that "was created to provide detailed mortality and population data to researchers, students, journalists, policy analysts, and others interested in the history of human longevity. The project began as an outgrowth of earlier projects in the Department of Demography at the University of California, Berkeley, USA, and at the Max Planck Institute for Demographic Research in Rostock, Germany "(Human Mortality Database 2014).

The data sets are the total (male and female) death rates. The three sets of data, which were analyzed, are Australia 2009, France 2012, and Switzerland 2011. All these countries have the high life expectancy (83.0, 82.3, and 82.8 respectively (Overall Life Expectancy 2012). The parameters of the death rate model (1) were estimated using the non-linear least squares (NLLSQ) regression procedure. The estimates of the parameters are displayed in Table 1, and the fitted death rates are shown in Figures 1, 2, and 3 below.

Table 1. Estimates of distribution (1) parameters

Country, Year	Estimates of distribution parameters			Proportion of variance of $h(t)$ explained by model, R^2
	a	$\ln(b)$	β	
Australia, 2009	186.3	47.44	10.86	0.98
France, 2012	228.8	69.77	14.58	0.96
Switzerland, 2011	204.8	62.01	13.51	0.96

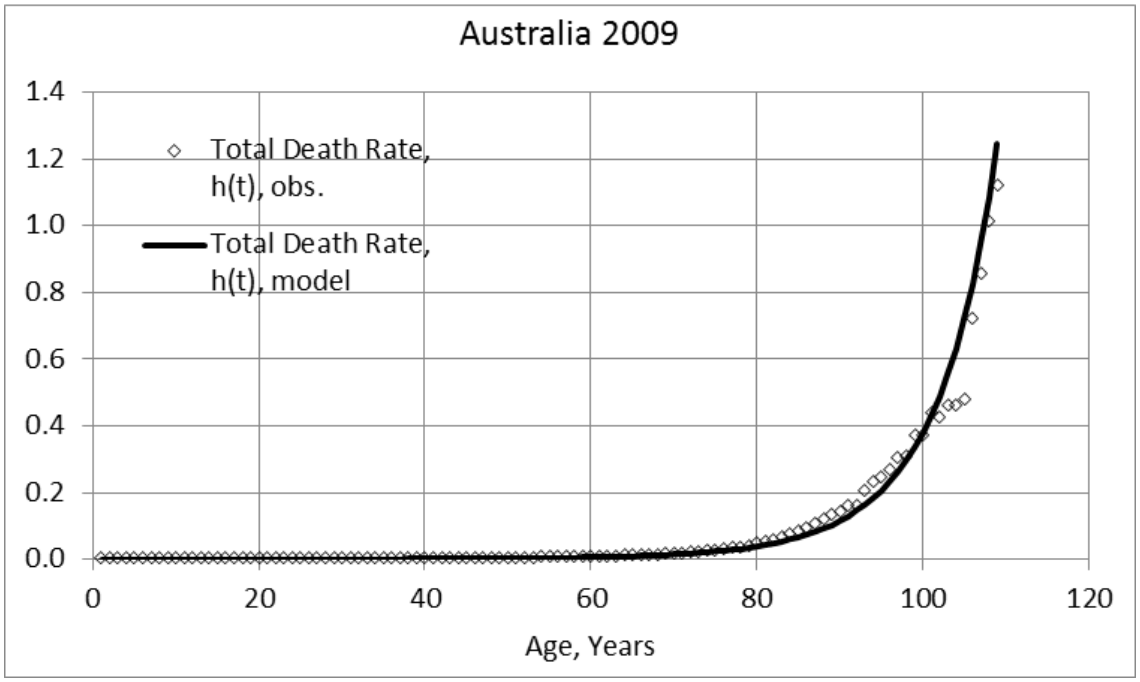

Figure 1. Total (male and female) death rates, Australia, 2009

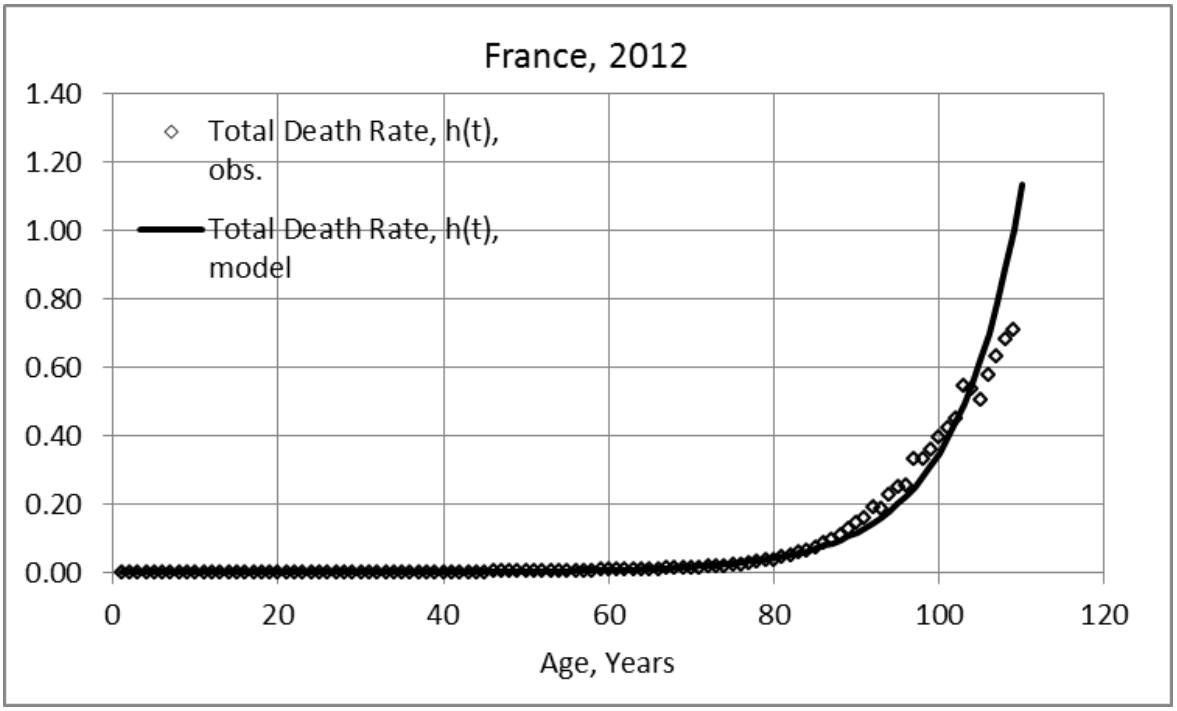

Figure 2. Total (male and female) death rates, France, 2012

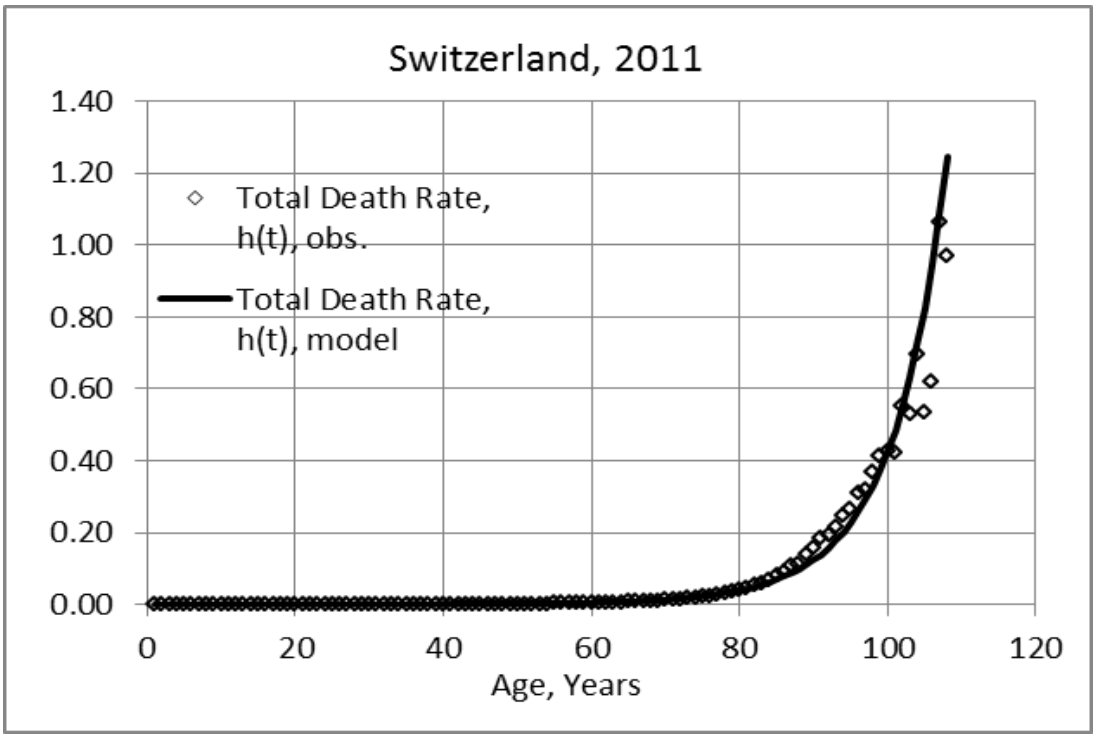

Figure 3. Total (male and female) death rates, Switzerland, 2011

Table 2. The two-sided 95% confidence limits on the model (1) parameters for the Australia 2009 data set

	Estimate	Lower Confidence Limit	Upper Confidence Limit
ln(b)	47.4434	38.2577	56.6292
β	10.8602	9.2662	12.454
a	186.3213	169.073	203.57

Table 3. Australia 2009. Correlation matrix of estimates

	b	β	a
b	1.000000	0.999853	0.996621
β	0.999853	1.000000	0.995114
a	0.996621	0.995114	1.000000

RESULTS AND DISCUSSION

With all the high proportions of variance explained by the model (R^2) for the data sets considered, the model fit could not be called perfect. The confidence intervals on the lifespan distribution parameters are rather wide, as revealed by Table 2, and the estimates of the parameters are strongly correlated as illustrated by Table 3. This is, to an extent, a typical situation, when the NLLSQ method is applied. On the other hand, one can notice that for all three data sets, the data points are above the fitted curve in the age interval of 80 – 100 years, and they are under the fitted model for the ages older 100 years. This model misfit for the older ages might be indication that a better lifespan distribution should be a competing risk distribution (Kaminskiy 2012). According to L. Hayflick (2000), "More than 75% of all human deaths in developed countries now occur in those

over the age of 75. If the causes of these deaths are resolved we will not become immortal but we will have revealed how death occurs in the absence of disease."

Based on this, we can assume that more adequate lifespan distribution will have two competing risks – the risk of death from diseases, and the risk of death from "pure" aging as "inexorable loss of physiological capacity in the cells of vital organs. . ." (Hayflick 2000).

ACKNOWLEDGMENT
The author is most grateful to the Human Mortality Database (http://www.mortality.org) allowing free access to the mortality data.

REFERENCES

1. Tan, Q. et al. (2004) Multidisciplinary Approaches in Genetic Studies of Human Aging and Longevity, Current Genomics, 5: 409–416
2. Hayflick, L. (2000) The future of aging. Nature 408: 267–269.
3. Weon, B. M. and Je, J.H. (2009) Theoretical Estimation of Maximum Human Lifespan, Biogerontology 10: 65–71
4. Ouellette,Q. and Bourbeau, R. (2011) Changes in the age-at-death distribution in four low mortality countries: A nonparametric approach, Demographic Research v. 25, article 19, 595–628
5. Human Mortality Database (2014), http://www.mortality.org
6. Overall Life Expectancy. List by the World Health Organization (2012)
7. Kaminskiy, M. (2012) Reliability Models for Engineers and Scientists, CRC Press