Single parameter of inverse proportion between mortality and age could determine all mortality indicators in the first year of life

Josef Dolejs *

University of Hradec Králové, Rokitanského 62, Department of Informatics and Quantitative Methods, 500 03 Hradec Králové, Czech Republic

H I G H L I G H T S

- Age trajectories of mortality are studied in 20 age categories in the first year and basic models are tested.
- All models except for the linear model and the linear model with the slope -1 in the log–log scale are rejected.
- All mortality indicators describing the first year could be determined by the parameter.

A R T I C L E   I N F O

A B S T R A C T

Mortality increase with age in adult population has been studied and modeled by many authors, but relatively little attention has been given to mortality decrease with age after birth. Data split in more detailed age categories can newly test mortality decrease with age. Age trajectories of mortality are studied in 20 age categories in the specific age interval 1–365 days. Four basic models mentioned in literature are tested here. The linear model and the linear model with the specific slope -1 in the log–log scale represent the most successful formalism. Mortality indicators describing the first year could be determined by a single parameter of the model with slope -1 in the log–log scale. All conclusions are based on published data which are presented as a supplement.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The relationship between mortality and age is investigated mostly for ages when mortality increases with age (Gompertz, 1825; Vaupel et al., 1998; Bongaarts, 2005; Patterson, 2010; Arbeev et al., 2011; Bebbington et al., 2011; Zheng et al., 2011 and many other studies). Relatively little attention has been given to mortality decrease with age after birth and four models found in literature are tested here. A smaller portion of the population dies after birth, but these deaths are significant when calculating life expectancy at birth. Generally, mortality is very high immediately after birth and it decreases with age faster than it increases for adults. Even the first look at age mortality trajectories shows that the age interval between the first day and the age of mortality minimum is shorter than the age interval between mortality minimum and the age when mortality reaches the same level in adult age. Mortality reaches the level of the first day of life after

the age of 60 years. For example, it is illustrated in Fig. 1 for both sexes in the Czech Republic in 2011.

What type of model could describe the mortality decrease after birth? Is it exponential, power or some other type of mathematical function? The historical list of studies on this topic is relatively short. The exponential Gompertz model used by Thiele was the first attempt (Thiele, 1871). Bourgeois Pichat composed the second important model describing the decrease of death in the age of n months was defined for the age interval 28–365 days (Bourgeois Pichat, 1946, 1951; Carnes et al., 1996; Knodel and Kintner, 1977). Bourgeois Pichat formula for cumulative deaths $q(n)$ up n months is:

$$q(n) = a + b\left[\ln(n+1)\right]^3 \quad \text{for 1 month} \leq n \leq 12 \text{ months} \tag{1}$$

A model of age trajectory of mortality for all age categories was designed later by Heligman and Pollard (1980) and Preston et al. (2001). Only the following term of the general formula describes mortality decrease after birth:

$$\mu(x) = A^{(x+B)^C} \quad \text{for } 0 < A < 1, 0 < B < 1, 0 < C < 1 \tag{2}$$

* Tel.: +420 776 703 629.
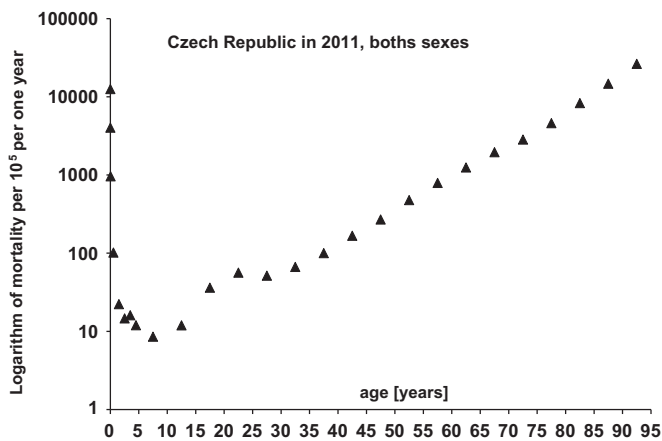E-mail address: djosef@post.cz

Fig. 1. Age trajectory of mortality in the semi-logarithmic scale for both sexes in the Czech Republic in 2011.

Generally, the simple linear model with two parameters in the log–log scale could be used and it formally corresponds to the Weibull distribution. For example, Bebbington et al. (2007a, 2007b) used the reduced additive and flexible Weibull distributions with the separation of mortality to exogenous or endogenous causes over the entire life span. Furthermore, the inverse proportion relationship between mortality and age was used to fit age trajectory of all diseases mortality (it means total mortality without non-biological causes; Dolejs, 2003, 2014).

$$\mu(x) = \mu_1/x \quad \text{or in the log} - \text{log scale} \quad \ln[\mu(x)] = \ln(\mu_1) - \ln(x) \quad (3)$$

2. Materials and methods

The numbers of deaths in specific age categories are extracted from Demographic Yearbooks of the Czech Republic published for individual years 2003–2011 (Czech Statistical Office, 2014). The numbers of deaths are obtained from table labeled there as: "F.04 Deaths: by age, year of birth, sex and marital status". The numbers of live birth are obtained from table labeled there as: "D.04 Live births: by age of mother, sex, legitimacy and birth order". The first table uses the following 21 age intervals: [0, 1) day, [1, 2) days, [2, 3) days, [3, 4) days, [4, 5) days, [5, 6) days, [6, 7) days, [7, 14) days, [14, 21) days, [21, 28) days, [1, 2) months, [2, 3) months … [11, 12) months. Two specific numbers of deaths are published in this table in each age category over the first month. The first number represents the number of death in specific age category for individuals born in the previous calendar year while the second number is the number of deaths born in the actual year. For example, an individual born on the tenth of November, 2010 who died on January 8th, 2011 could be found in the Yearbook for 2011 in the age category 1–2 month in the subset labeled as 2010. The same age category contains the subset labeled as 2011 with the individuals born in 2011.

The first age category labeled as "the First day" could be in practice interpreted by two ways:

a) as the number of deaths which died during the first 24 h or
b) as the number of deaths which died during the first calendar day of life.

For this uncertainty, the age category "the First day" is not used here.

Remaining 20 age categories are used while the age interval 11–12 month represents the last age category. Furthermore, numbers of deaths are divided according to sex into the three

categories: "both sexes", "males" and "females". All data are in the file "Data.xls" presented as a supplement.

3. Results

Cohorts with any age category containing no deaths are excluded, with the remaining 20 out of 27 cohorts used in further analysis. If the mortality axis is on the logarithmic scale, and age is not on the logarithmic scale, then age trajectories of mortality show a concave relationship in all 20 cases (the linear relationship in the projection corresponds to the exponential model). On the other hand, an approximate linear relationship is observed in the log–log scale in all cases.

In addition to the models (1)–(3) mentioned above, the simple exponential Gompertz model (Thiele, 1871) expressed here by (4) and also simple power model with two parameters are tested.

$$\mu(x) = e^{(Ax+B)} \quad \text{or for logarithm of mortality} \quad \ln[\mu(x)] = Ax + B \quad (4)$$

Estimations of parameters are calculated using the standard method of least squares (LS) for all models in the age interval the 1st–365th days, and residuals are investigated. The residuals calculated in the three models (1), (2) and (4) do not match the assumptions of the regression model. They are U-shaped in the model (4) in all cases and they are strongly dependent on age and also on the logarithm of age in the two models (1) and (2) in all cases ($p < 0.001$). The residuals calculated in the four models (1)–(4) are shown as an example in Figs. A1–A4 in the "Appendix A" for cohort born in 2011. Data of the cohort born in 2011 with the trajectories calculated in the four models are shown in Fig. 2.

The same findings are confirmed in all cohorts and, consequently, the three models (1), (2) and (4) are found to be unsuitable. On the other hand, the residuals calculated in the linear model are random in the log–log scale and independent of logarithm of age (the linear model in the log–log scale corresponds to the simple power model with two parameters). A p value of less than 0.05 is considered significant in all cases. At first, the linearity is tested using LS in the log–log scale and parameters in the following quadratic model (5) are calculated using LS.

$$\ln[\mu(x)] = \text{constant} + \gamma.\ln(x) + \delta.\ln(x).\ln(x) \quad (5)$$

The residuals calculated in the model are random, they are not U-shaped and they are not dependent of logarithm of age in all cohorts. The residuals calculated in the model (5) for cohort born in 2011 are shown in Fig. A5 as an example in the "Appendix A". Parameter δ is significant only in two from 20 cases. More specifically, quadratic model (5) is not rejected for both sexes born in
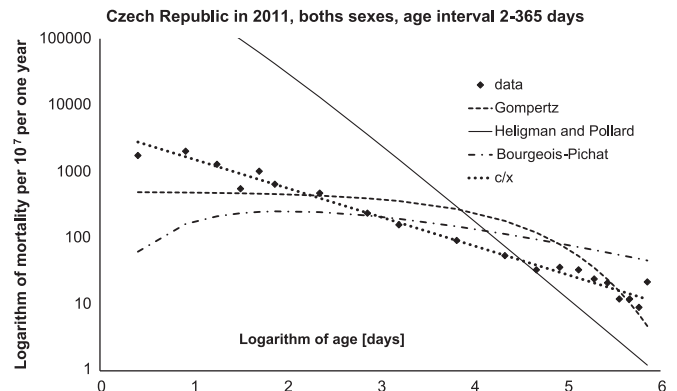

Fig. 2. Data of the cohort born in 2011 with the age trajectories of mortality calculated in the four models using LS in the log–log scale.