

METHOD OF CONSTRUCTING LIFE TABLES

Basic methodological notes

The complete life tables are calculated from the third main group of demographic events (age-period specific). Input death probabilities are computed indirectly, which means they are derived from real (observed) age-specific mortality rates (see below). The life tables are detailed i.e. by single year of age with an open age interval for 105+. They are computed separately for males and females.

To exclude random fluctuations, the life tables for regions (NUTS 3) were calculated for a two-year period and the life tables for the districts and Administrative Districts of Municipalities with Extended Powers (AD MEP) for a five-year period. Nevertheless, low numbers of deaths and used (in time and spatially uniform) methodology for smoothing and modeling the observed mortality rate can stand behind the fluctuations in time series.

The basic input data are the data on the number of deaths and mid-year population by age and sex in the given year, which has the CZSO at its disposal and uses in its outputs. Only the input data available for both deaths and population at single age detail enter into the calculation (for example in the years 1920–1924 the input data were used data only for ages of 0–94 years included). The input data always refers to the regional breakdown valid on 1 January of the given year of the event.

The Czech Statistical Office has used this described methodology of computing life tables since 2018 (with the beginning of data processing for the year 2017). Mainly for monitoring of dynamics of mortality development, in autumn 2018 the CZSO prepared a retrospective set of life tables in time series calculated uniformly by using the new methodology; the life tables for the Czech Republic cover the period starting 1920, in the case of life tables for cohesion regions, regions and districts, the retrospection concerns the period starting 2001. This retrospective set of life tables was published in the publication *Life Tables in Time Series* (CZSO, 2018).

Life tables indicators

The **number of deaths (D_x)** states the number of deaths by age (x) during the reference period.

The **population (P_x)** states the mid-year population by age (x) or the sum of mid-year populations in the case of multiple-year periods. In life tables for the Czech Republic, cohesion regions (NUTS 2), regions (NUTS 3) and districts the mid-year population is defined as the balanced population on 1 July of a given year. In life tables for the administrative districts of MEP the mid-year population is defined as an average of the start-period and end-period population of a given year.

The **death probability (q_x)** expresses the probability that an individual at the exact age of x years will die in a given period, i.e. before the exact age of $x+1$ years:

$$q_x = \frac{m_x}{1 + (1 - a_x) \cdot m_x}$$

where m_x is the mortality rate at a given age and the parameter a_x is the average number of years lived within the age interval $[x, x+1)$ for people dying at that age (see below).

The **table number of survivors (l_x)** is a hypothetical number of individuals alive at the exact age of x years out of 100,000 live births (table radix $l_0 = 100,000$), given the mortality conditions of the reference period:

$$l_{x+1} = l_x \cdot (1 - q_x) \quad .$$

The table number of deaths (d_x) is a hypothetical number of individuals who die at the complete age of x years; it is computed as the difference between two subsequent table numbers of survivors:

$$d_x = l_x - l_{x+1} .$$

The **table number of person-years (L_x)** is a hypothetical number of person-years lived by the life-table population in the age interval $[x, x+1)$:

$$L_x = l_x - (1 - a_x) \cdot d_x .$$

The **auxiliary indicator (T_x)** expresses the number of years of life to be lived by the life-table population (not by an individual) at a given age x . It is the accumulation of L_x from the age of x to the highest age of the table.

$$T_x = \sum_x^{105+} L_x$$

The **life expectancy (e_x)** shows the average number of years an x -year-old individual can expect to live, given the mortality conditions of the reference period. It is a synthetic indicator reflecting mortality conditions in all age groups in a given year.

$$e_x = \frac{T_x}{l_x}$$

Detailed description of constructing complete life tables

A. Input data processing

1. Calculation of real (observed) mortality rates

The so-called real (observed) mortality rate (m_x) is computed from empirical data as the ratio of the number of deaths and the mid-year population at a given age, sex, and area:

$${}_t m_x^{\text{CR, cohesion region}} = \frac{{}_t D_x}{1.7.(t) P_x}$$

$${}_{t-1,t} m_x^{\text{region}} = \frac{{}_{t-1} D_x + {}_t D_x}{1.7.(t-1) P_x + 1.7.(t) P_x}$$

$${}_t m_x^{\text{district}} = \frac{\sum_t {}_t D_x}{\sum_t 1.7.(t) P_x}$$

$${}_t m_x^{\text{AD MEP}} = \frac{\sum_t {}_t D_x}{\sum_t \left(\frac{1.1.(t) P_x + 31.12.(t) P_x}{2} \right)}$$

2. Smoothing of real (observed) mortality rates

To treat possible random fluctuations in real (observed) mortality rates, which are common particularly at ages with low mortality or the low number of survivors, the mortality rates are smoothed by the generalised additive model (the GAM method) combined with the so-called P-splines; this method is sometimes called the P-GAM method (Eilers and Marx, 1996). The age is the predictor, the number of deaths is the predicted figure, and the population is used as an offset. Poisson distribution of the number of deaths is assumed. The level of smoothing is adapted to the number of deaths at a given age; it means that a lower number of deaths is smoothed more than a higher number of deaths; the significance of the original phenomenon/trend is thus maintained.

3. Mortality modeling at high ages

Given the low numbers of deaths at high ages and lower reliability of data on the mid-year oldest population, the smoothed observed mortality rates are substituted by model figures at the oldest old ages. Generally, mortality models work with a continuous mortality rate function, the so-called force of mortality, denoted by μ_x . In practice, it is assumed, based on the monotonic increase of mortality with age that the force of mortality and the mortality rate at a given age are in this connection approximately (Thatcher et al., 1998):

$$m_x \cong \mu_{x+0.5} \quad .$$

The model selected here is based on a logistic curve, which takes into consideration the deceleration in mortality increase with age. Specifically, the model named by Kannisto is used (Thatcher et al., 1998). So the programme models the age-specific mortality by the following function:

$$\mu_x = \frac{\alpha e^{\beta x}}{1 + \alpha e^{\beta x}} + c$$

where x is the age, α and β are parameters, and c is a constant. The parameters are estimated by the maximum likelihood estimation method (MLE). The default parameters estimate is based on an assumed Poisson distribution of the number of deaths (D) at the age (x). Based on these assumptions, the logarithm of the likelihood function (L) is maximized as:

$$\ln L = \sum_x D_x \ln(\mu_x) - E_x \mu_x$$

During the optimization process (of the negative logarithm) of the likelihood function the programme works with its default values of parameters. For estimation of the parameters, the age range of 70–90 years is set.

As regards a transition from smoothed data to model data, it is assumed that the age, from which model data characterize mortality better than smoothed data is higher than 75. For smooth progress of the substitution of smoothed data with model ones, the age denoted by y (higher than 75 years) is sought first, where the difference between the smoothed and the model figure is minimal:

$$|m_x^{\text{smooth}} - m_x^{\text{model}}| = \min, \quad y \geq 75$$

A step further, a gradual change from the smoothed figures to model ones is made for the ages $x \in [(y - 4), (y + 4)]$ according to:

$$m_x^{\text{grad}} = \left(1 - \frac{x - y + 5}{10}\right) \cdot m_x^{\text{smooth}} + \left(\frac{x - y + 5}{10}\right) \cdot m_x^{\text{model}}$$

4. For the computation of life tables the following data on age-specific mortality rates are used as the input:

at the age of 0 years	real (observed) figures,
at the age of 1 to $(y - 5)$ years	smoothed figures,
at the age of $(y - 4)$ to $(y + 4)$ years	graduation from smoothed to model figures,
at the age of $(y + 5)$ years and over	model figures.

B. Conversion of real smoothed data to table figures

The basis for the calculation of life tables in an indirect way is to convert age-specific mortality rates (m_x) to death probabilities (q_x). Based on the assumption that the real mortality rates and the table ones are equal and on the assumption that mortality rates in a hypothetical table cohort are constant, the following conversion formula is used (Chiang, 1984):

$$q_x = \frac{m_x}{1 + (1 - a_x) \cdot m_x} \quad .$$

The parameter a_x quantifies the average number of person-years lived in the given age interval by those who die in the interval. With regards to the generally accepted assumption of a uniform distribution of the exact age of deceased persons, the value is set to 0.5 for all ages except 0 and the last (open) age interval.

In the case of the estimation of the average age of deceased infants (a_0) the most exact way of calculation has been used since input data of 1986; it is computed from the real (observed) individual data as the observed proportion of the mean age at death for children who died in their first year of life (expressed in days) in the length of the whole calendar year (Andreev and Kingkade, 2015):

$$a_0^{\text{poz}} = \left(\frac{\sum X_{D_{i,0}}^{(\text{days})}}{\sum D_{i,0}} \right) / 365$$

where $D_{i,0}$ is an individual death in the first year of life (infant period) and $X_{D_{i,0}}^{(\text{days})}$ is an individual age of the deceased infant in days. When processing life tables for the Czech Republic for the years 1920–1985, the parameter a_0 was set as default value 0.1 due to insufficient detail of the available infant death data.

In the case of lower regional breakdowns, the value is always the one calculated from data for the whole Czech Republic in the given period.

Concerning the last open age interval (105+ here) the parameter a_0 is calculated as (see for example Preston et al., 2000):

$$a_{105+} = \frac{1}{m_{105+}}$$

The next steps of life tables calculation follow the computed q_x values (see Life tables indicators).