Database of age trajectories of mortality in 110 countries and web application: Data report

Josef Dolejs*

Department of Informatics and Quantitative Methods, University of Hradec Králové, Hradec Králové, Czechia

Introduction

The mortality rate is an important indicator of population health. It shows a decline in age after birth and increases with age in older age groups. The increase in adults is approximately exponential and may be interpreted as a manifestation of aging that affects all individuals (1–14).

From an analytical perspective, age remains a deterministic variable in the relationship between age and mortality, and coefficients of determination are higher than 0.99 (4, 10–13). Changes in the mortality rate with age during childhood are quicker, and the steep decline in mortality is accompanied by age-based changes in predominant causes of death (15–20). Congenital anomalies or impairments originating in the perinatal period result in 85% of all deaths during the first 4 weeks of life. However, after the age of 5 years, their contribution reduces to less than 10% (17–20). As such, age is very reliable data as compared to the cause of death.

The author presents the database and web application following the previous studies that included 14 European countries and 25 countries from the four continents (19, 20).

Age trajectories of mortality (ATMs) due to some diseases after birth showed an important phenomenon that may be important in pediatrics. It may be verified or rejected in other regions and countries using the presented database. Age trajectories of total mortality (ATTMs) were the most important ATMs. The decrease in ATTMs was described by the model of the inverse proportion with coefficients of determination higher than 0.99 (15, 17, 19, 20).

The ATMs as a result of major disease groups were decreased with age after the first year of life, with one exception, malignant neoplasms (15, 17, 19, 20). The ATM due to neoplasms was declined only during the first month of life and remained age-independent from the first month till the age of 20 years.

Further, ATMs due to major disease groups (e.g., "neoplasms" or "infectious diseases") were composed of a larger spectrum of diseases. The database enables researchers to investigate the composition of ATM (the spectrum of diseases that contributed to some ATM). Researchers or other users may combine diseases in a selected population, and the resulting ATM will be transferred to the web application. The results shown in the previous studies may be verified or rejected using the database in 110 countries and in 10 aggregated populations.

Methods

World Health Organization raw data

The WHO Raw Mortality database was the main source of data (21). It contains the number of deaths within specific age categories in different countries (21).

The cause of death was determined using the specific revision of the International Classification of Diseases (ICDs) in the database. The database contains the 7th–10th revisions of ICD. Specific revisions were used by different countries at different time periods. The use of the 10th revision of ICD (ICD10) was the first main criterion to segregate data (22). The number of deaths due to causes classified as per ICD10 was found in two files in 2019 (21).

The database does use different categories for the age that differ mainly during the first year of life. The following four age categories in the first year of life were the most detailed: 0–24 h after birth, 1–7, 7–28, and 28–365 days. Not all countries use these four categories, and hence, the second criterion for segregating data was the four age categories.

Thus, data using ICD10 were selected, and the resulting two files "Morticd10_part1.txt" and "Morticd10_part2.txt" were aggregated to create one file. Within this, the rows with four age categories in the first year were used as filters for segregation. Figure 1 is a diagrammatic representation of this process.

Applying the ICD10 and age criteria reduced the number of countries to 122. The number of calendar years in some countries were reduced too. The WHO database uses the four mentioned age categories in the first year of life, a single age category for the age range 1–5 years and ages above 5 years are grouped with an interval of 5 years, each up to ages 90–94 years.

The next important step was sourcing data on the number of people alive. WHO database contains the number of living people in the file "pop.txt" but it uses one age category in the age range 1–5 years in some countries in some years. For this reason, the set of years does not reach the set of years used in the two files "Morticd10_part1.txt" and "Morticd10_part2.txt". For example, only the years 1966–1989 may be found in the US with single-year age categories. The "United States Census Bureau" database of living people in a single-year age group may solve the problem, and it was used to automate the process of ATM construction (23). A total of 110 out of 122 countries were available in the database, and thus, the ATM was constructed for these 110 countries.
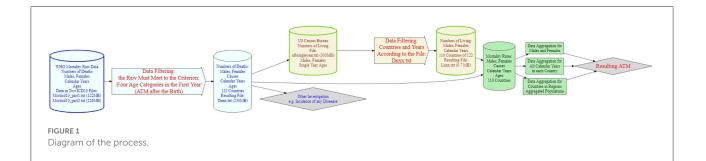
Further, 10 large, aggregated populations that contained countries from different regions of the world were created. The largest aggregated population "P110" contained all the 110 countries. The number of people alive per 1 year in "P110" was approximately 2.286 billion, and it represented about one-third of the whole world population during the study period. The list of 110 countries may be shown in the web application

if an aggregated population is selected in the panels: "ATM2," "CACNS," and "Main Chapters."
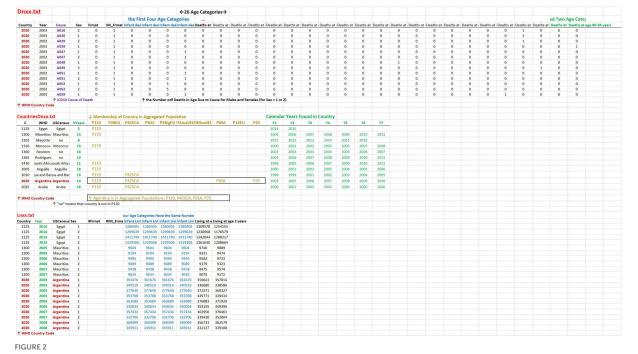
Number of people alive—United States Census Bureau

The database "United States Census Bureau" enabled us to download file "idbsingleyear.txt" that included the number of people alive (men and women) in 110 countries, in 1-year age categories, and in calendar years (23). The set of 122 countries and the set of calendar years in each country were used to find the suitable number of people alive (men and women) in the file "idbsingleyear.txt." The process reduced the number of countries from 122 to 110. Three regions, "*United Kingdom England and Wale*s," "*United Kingdom, Northern Ireland*," and "*United Kingdom, Scotland*," were not used separately as all three regions were referred to as "*United Kingdom*" in the file "idbsingleyear.txt." Further, nine countries were not found in the file "idbsingleyear.txt," and they were excluded from the list of 122 countries: "*Mayotte*," "*Reunion*," "*Rodrigues*," "*French Guiana*," "*Guadeloupe*," "*Martinique*," "*Netherlands Antilles*," "*Occupied Palestinian Territory*," and "*Serbia and Montenegro*" (Figure 1). The number of people alive in the first age category was used to calculate the ATM in the four age categories as categorized by WHO for the first year of life. The next four single-year age categories were used to calculate ATM in the age range 1–5 years. Similarly, the 5-year age categories (from ages 5 to 95) were calculated from the single-year age categories.

Halley method—Aggregation of calendar years and countries

The 1-year period is a typical time unit used in epidemiology, biology, and social sciences. Biological events (such as death or disease) in different ages may be studied in one specific calendar year across age categories (cross-sectional description). The other possibility is the longitudinal observation of individuals born in the same year (generation study or longitudinal study). If the time unit "1 year" was replaced with the time unit "1 day" (e.g., for some biological, social, or cultural reasons), events observed for 1 year represented the aggregation of the events that occurred in 365 days. The aggregation of more calendar years had a similar meaning. Age was assumed as the main factor, and all other factors were assumed to be less significant. Furthermore, the existence of general mechanisms was assumed, as demonstrated in the ATM after birth. For these reasons, ATM may be constructed in as large a population as possible (19, 20). Including more regions and more calendar years within the analysis may eliminate all factors other than age, rendering the impact of age more visible. The following standard definition of

FIGURE 1
Diagram of the process.

FIGURE 2
Structure of three main files. The structure of the two files "Dnxx.txt" and "Lnxx.txt" is similar. Each row contains the numbers of deaths in "Dnxx.txt" and the numbers living in "Lnxx.txt," in a specific calendar year, for men and women separately in age categories. The first two columns contain country codes and calendar years in both files. The third column contains the cause of death in "Dnxx.txt" while the country name in "Lnxx.txt." The other columns have the same meaning in the two files. The fourth column contains sex, the next two columns contain an index of WHO format of age categories, and the last columns represent 26 age categories (columns: 7, 8, …32). The file "ContriesDnxx.txt" contains information about a specific country. The columns contain the WHO code, the name of the country used in WHO, the name of the country used in the US Census Bureau, and the number of calendar years. Columns 5–14 contain the indication if the country was used in aggregated population (a specific column corresponds to a specific aggregated population). The last columns contain the list of calendar years found in the specific country in the WHO database.

the force of mortality at age x was used:

$$\mu\left(x\right) = \lim_{h\to 0} \frac{D\left(x+h\right)}{L\left(x\right)} = -\frac{\frac{dS(x)}{dx}}{S(x)} \cong \frac{Di}{Li}\cdot\frac{1}{\left(Bi - Ai\right)} \quad (1)$$

where $D(x+h)$ is the number of deaths in a small age range, [x, x+h), the infinitesimal increment h is positive. L(x) is the number of living people at the age x, $S(x)$ is the survival function (percentage of living people at age x), which is valid in principle: $S(x) = 1 - F(x)$, where $F(x)$ is the cumulative distribution

function of the probability of death. The empirical value Di is the number of deaths within an age range $[Ai, Bi]$, while Li is the size of the population among which the deaths occurred. Empirically, changes in Li within an age interval $[Ai, Bi]$ were very small in childhood, when compared with changes in Di, and the number of living people in Li within an age interval $[Ai, Bi]$ was used instead of the average number of living people in the region and calendar period. Namely, population Li went through the "window" in time, and it was the meaning of the product $Li\bullet(Bi - Ai)$ in Equation (1). Element Bi – Ai differed from 1

only in the first four age categories where the four values, 1/365 years; 6/365 years; 21/365 years; and (365–28)/365 years, were used. It equaled to 1 in the 5-year age categories because Di and Li were detected in one calendar year. The unit corresponding to the mortality rate was "person-years," which meant the number of years lived by members of the population between the ages A_i and B_i. Uncertainty or possible demographic error of Li was discussed in the previous studies in detail and they were negligible with the respect to resulting ATM (11–13, 17–20).

The ATMs were assumed to be unknown theoretical curve and was constructed using the right side of the equation (1). Mortality rates in different age groups described the group's Li in the same way a decay constant may describe the force of a radioactive decay on different radionuclides (the different radionuclides correspond to the group's Li in the parallel) (19, 20, 24).

Zero deaths in a given age group may be the major problem for the construction of ATM from a specific disease or from a set of diseases. While zero deaths due to some disease may occur within a specific age category, in a specific calendar year, at least one death may occur within the same age category in other calendar years. The inclusion of additional calendar years may remove this obstacle. The method was first utilized by the well-known astronomer and mathematician Edmond Halley (25). This method enabled the calculation of the mortality rate within one age category based on the number of deaths and living persons in several years (25–27). Standard epidemiologic and demographic interpretations of the specific value of mortality rate within a specific age category differed from the values calculated for one calendar year. However, the interpretation was the same for all age categories, and the resulting ATM may be interpreted as a hallmark of aging. In addition, including more regions and calendar years may eliminate factors other than age.

The ATM in one country in one calendar year may be constructed separately for men and women using three main data files, which are described in the paragraph "Three Main Data Files."

Aggregated populations

In total, 10 aggregated populations were mainly created to eliminate zero cases due to specific diseases. The aggregated population may be selected by the user in the web application in the panels "ATM2," "CACNS," and "Main Chapters." A list of countries in the aggregated population is automatically shown in the panels.

The largest aggregated population "*P110*" contained all 110 countries. "*P28EU*" contained 28 European countries, "*P42SCA*" contained 42 countries from South and Central America, and "*P6AS*" contained six Asian countries. Three groups of European countries from the previous study were used in the aggregated populations: "*P5BigEU*," "*P5SmallEU*," and "*P4NordEU*." (19) Aggregated population "*P6SA*" contained six countries of South America (20). In addition, two aggregated populations "*P14*" and "*P25*" used similar to their use in the previous studies ("*P14*" = "*P5BigEU*" + "*P5SmallEU*" + "*P4NordEU*" and "*P25*" = "*P14*" + "*P6SA*") (19, 20).

Calendar years used by a specific country in the database present in the web application if a specific country is selected in the panels "ATM2," "CACNS," and "Main Chapters" (the list found in the file as "CountriesDnxx.txt").

Three main data files

The three main data files, "Dnxx.txt," "CountriesDnxx.txt," and "Lnxx.txt," may be downloaded in the web application in the panel "Introduction." The structure of the files is shown in Figure 2.

The data may be obtained in different ways than in the web application. For example, in the aggregation of calendar years, men and women may be ignored. In this case, ATM in the specific calendar year may be constructed using the files "Dnxx.txt" and "Lnxx.txt" (e.g., in Excel). Calendar years and age categories were unified in "Dnxx.txt" and "Lnxx.txt." Users may calculate ATM using the formula (1) (the element Bi-Ai differs from the value of one in the first four age categories).

Data investigation and web application

There are more possibilities on how to run the web application:

(a) Directly on the Internet: https://atm-dolejs-josef.shinyapps.io/ATTM/
(b) Download the directory from the Supplementary Material
(c) Download the directory here https://lide.uhk.cz/fim/ucitel/dolejjo1/

If the directory is downloaded, the application may run according to the instructions in the file "Run Application in Personal Device with Windows.pdf."

The application contains 10 panels: Introduction, ATTM1, ATTM2, CACNS, Main Chapters, Detail Codes1, Detail Codes2, Combinations of Main Chapters, Significance of Main Chapters in Age, and Other Models.

The first panel "Introduction" contains an introduction to ATM after the birth and a short explanation of the theory of congenital individual risks of death (TCIR) (16, 17, 19, 20). The panel contains three keys to download three main

files "Dnxx.txt," "DnxxCountries.txt," and "Lnxx.txt," the fourth button to download the file "TCIR.pdf" with the explanation of TCIR, and fifth key to download the file "Numerical_Bending_ATM_in_TCIR.xlsx" with the demonstration of bending ATM according to TCIR (16, 17, 19, 20).

The second panel "ATTM1" contains an illustrative animation of ATTM in 110 countries and in 10 aggregated populations. The panel "ATTM2" contains the same plot, while each population may be individually selected. Moreover, a list of all countries and the mean number of people alive in all age categories in specific calendar years in the selected aggregated population is shown. The two identical plots in ATTM1 and ATTM2 contain coefficient of determination calculated in the model of the inverse proportion in the age range 0–10 years (in the first nine points). Confidence bands are common for all populations. They were calculated as limits between which 95% of empirical values, in each age category, were detected in the 110 countries.

The panel "CACNS" is similar to the panel "ATTM2"; however, ATM due to congenital anomalies of the central nervous system (CACNS) is shown. The age range used in the model of the inverse proportion with the 95% confidence bands is 0–40 years.

In the panel "Main Chapters," one aggregated population and one main chapter of ICD10 (one of 22) may be selected. The ATM plot using the selection of the main chapter in the chosen aggregated population with four illustrative straight lines is displayed. Two straight lines have the slope −1 and go exactly through to the second point (the first week) and through to the fifth point (1–2 years), respectively. One straight line has the slope −2 and goes through the fifth point. The last straight line has a slope 0 and corresponds to the age independence. These straight lines may be important and may be used to roughly classify ATM (the straight lines are not the regression lines but they have an exact slope and go through observed mortality rates in specific age categories). In addition, the spectrum of ATMs that are significant to ATTMs are shown in the panel below in two plots, in two age ranges "0–10 years" and "1–10 years."

The aggregated population and main chapter of ICD10 selected in the panel "Main Chapters" are used automatically in the panel "Detail Codes1." The resulting list of all disease codes is displayed. The users may select their own set of codes in the section "Choose Diseases of ICD10" and create a specific group of diseases. The ATM due to the group of diseases is displayed with four indicative straight lines (the same four straight lines as described in the panel "Main Chapters"). There are two keys to download data files with the list of selected codes and with the ATM data as per the disease groups.

Three illustrative examples of the use of the panel "Detail Codes1":

(a) It was shown that ATMs due to the second chapter of ICD10 (Neoplasms) were age independent in the age range from the first month to 25 years (16, 17, 19, 20). If the aggregated population is chosen in the panel "Main Chapters," the list of all ICD10 codes used in all age categories will be seen. Each code may be selected by a click on the left panel (it may be removed by a click on the right panel). The ATM due to the selected set of codes is calculated, and a plot is shown (two possible levels of ICD10 were used in the WHO database, and the problem is described below).

(b) It was shown that ATMs due to the first chapter of ICD10 (and due to other chapters) were age independent or declined slowly with age in the first year of life. Simultaneously, the decrease with age was faster after the first year (usually with the slope −1) (16, 17, 19, 20). The bending ATMs were typical for many main chapters of ICD10 in which congenital impairment was not determined as the cause of death. It is possible to study the spectrum of diseases in the selected main chapter and in the selected aggregated population.

(c) It was shown that ATMs due to CACNS (Q00–Q07, congenital malformations of the central nervous system) were inversely proportional to age in higher age categories (e.g., 40 years) (16–20). If the aggregated population and the 17th chapter of ICD10 (all congenital anomalies) are selected in the panel "Main Chapters," then the study similar to the previous two examples "a, b" may be accomplished (only codes Q00–Q07 may be selected).

Some basic models may be tested directly in the panel "Detail Codes1," and ATM may be downloaded as a data file for other investigations.

The panel "Combinations of Main Chapters" enables to create any combination of main chapters of ICD10 in the selected aggregated population. For example, the contribution of the main chapters to ATTM may be studied here. Key "Downloaded Mortality Rates to CSV File" in the panel "Combination of Main Chapters" may be used to download mortality rates for the selected main chapter in the selected aggregated population.

Proportions of main chapters in selected aggregated population in age categories are shown in the panel "Significance of Main Chapters in Age." Primarily, it shows that the significance of cases with congenital impairments (the 16th and the 17th chapter of ICD10) dramatically decreases with age. The special group of diseases "Other Diseases" contains the first 15 chapters of ICD10, excluding the second chapter of ICD10 (Neoplasms) (16, 17, 19, 20).

Historical models found in the literature, which may describe ATTM after the birth, are displayed in the last panel "Other Models." (11–13).

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

Author contributions

The author confirms being the sole contributor of this work and has approved it for publication.

Funding

Acknowledgments

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpubh.2022.911589/full#supplementary-material

References

1. Gompertz B. On the nature of the function expressive of the law of human mortality. *Philos Trans R Soc London.* (1825) 115:513–85. doi: 10.1098/rstl.1825.0026

2. Makeham W. On the law of mortality and the construction of annuity tables. *J Inst Actuar.* (1860) 8:301–10. doi: 10.1017/S204616580000126X

3. Heligman L, Pollard JH. The age pattern of mortality. *J Inst Actuar.* (1980) 107:49–75. doi: 10.1017/S0020268100040257

4. Riggs JE. Longitudinal Gompertzian analysis of adult mortality in the US, 1900–1986. *Mech Ageing Dev.* (1992) 54:235–47. doi: 10.1016/0047-6374(90)90053-I

5. Willemse WJ, Koppelaar H. Knowledge elicitation of gompertz' law of mortality. *Scand Actuar J.* (2000) 2:168–79. doi: 10.1080/034612300750066845

6. Hayflick L, Moorhead PS. The serial cultivation of human diploid cell strains. *Exp Cell Res.* (1961) 25:585–621. doi: 10.1016/0014-4827(61)90192-6

7. Hayflick L. The limited in vitro lifetime of human diploid cell strains. *Exp Cell Res.* (1965) 37:614–36. doi: 10.1016/0014-4827(65)90211-9

8. Strehler BL, Mildvan AS. General theory of mortality and aging. *Science.* (1960) 132:14–21. doi: 10.1126/science.132.3418.14

9. Lin XS, Liu X. Markov aging process and phase type law of mortality. *North Am Actuarial J.* (2007) 11:92–109. doi: 10.1080/10920277.2007.10597486

10. Dolejs J. The extension of Gompertz law's validity. *Mech Ageing Dev.* (1998) 99:233–44. doi: 10.1016/S0047-6374(97)00104-8

11. Dolejs J, Maresova P. Onset of mortality increase with age and age trajectories of mortality from all diseases in the four nordic countries. *Clin Interv Aging.* (2017) 12:161–73. doi: 10.2147/CIA.S119327

12. Dolejs J. Age trajectories of mortality from all diseases in the six most populated countries of the south america during the last decades. *Bull Mathematic Biol.* (2014) 76:2144–74. doi: 10.1007/s11538-014-0005-0

13. Dolejs J. Modelling human mortality from all diseases in the five most populated countries of the European Union. *Bull Mathematic Biol.* (2017) 79:2558–98. doi: 10.1007/s11538-017-0341-y

14. Izsak J, Juhász-Nagy P. On the explanation of the changes in age of the concentration of the death causes. *Zeitschrift für Alternsforschung.* (1984) 39:31–6.

15. Dolejs J. Single parameter of inverse proportion between mortality and age could determine all mortality indicators in the first year of life. *J Theor Biol.* (2016) 397:193–8. doi: 10.1016/j.jtbi.2016.03.007

16. Dolejs J. Theory of the age dependence of mortality from congenital defects. *Mech Ageing Dev.* (2001) 122:1865–85. doi: 10.1016/S0047-6374(01)00325-6

17. Dolejs J. Analysis of mortality decline along with age and latent congenital defects. *Mech Ageing Dev.* (2003) 124:679–96. doi: 10.1016/S0047-6374(03)00063-0

18. Dolejs J, Homolkova H, Maresova P. Modeling the age-associated decrease in mortality rate for congenital anomalies of the central nervous system using WHO metadata from nine European countries. *Front Neurol.* (2018) 9:585. doi: 10.3389/fneur.2018.00585

19. Dolejs J, Homolkova H. Why does child mortality decrease with age? Modeling the age-associated decrease in mortality rate using WHO metadata from 14 European countries. *Front Pediatrics.* (2020) 8:2296–360. doi: 10.3389/fped.2020.527811

20. Dolejs J, Homolkova H. Why does child mortality decrease with age? Modeling the age-associated decrease in mortality rate using WHO metadata from 25 countries. *Front Pediatr.* (2021) 9:657298. doi: 10.3389/fped.2021. 657298

21. World Health Organization. Mortality ICD10 (2019). Available online at: https://www.who.int/data/data-collection-tools/who-mortality-database (accessed January 25, 2019).

22. World Health Organization. *The International Classification of Diseases*, 10th Revision, 3 digit codes (1997). Available online at: https://icd.who.int/browse10/2016/en#/ (accessed January 15, 2017).

23. Bureau of the Census (2019). Available online at: https://www.census.gov/programs-surveys/international-programs/about/idb.html (find the section "How to Access" and the text "You can also just download the entire dataset) (accessed May 3, 2019).

24. Dolejs J. Decrease of total activity with time at long distances from a nuclear accident or explosion. *Radiat Environ Biophys.* (2005) 44:41. doi: 10.1007/s00411-005-0276-9

25. Halley E. An estimate of the degrees of mortality of mankind, drawn from curious tables of the births and funerals at the city of Breslaw, with an attempt to ascertain the price of annuities on lives. *Philos Trans.* (1693) 17:596–610. doi: 10.1098/rstl.1693.0007

26. Bellhouse DR, A. new look at Halley's life table. *J R Statist Soc A.* (2011) 174:823–32. doi: 10.1111/j.1467-985X.2010.00684.x

27. Luy MA, Wittwer-Backofen U. The halley band for paleodemographic mortality analysis. In: Bocquet-Appel JP, editor. *Recent Advances in Palaeodemography.* Dordrecht: Springer (2008). p. 119–41.