CHARLES UNIVERSITY IN PRAGUE

Faculty of Science

Department of Demography and Geodemography

Study program: Demography

Mgr. Klára Hulíková Tesárková

SELECTED METHODS OF MORTALITY ANALYSIS FOCUSED ON ADULTS AND THE OLDEST AGE-GROUPS

Vybrané způsoby zkoumání procesu úmrtnosti se zaměřením na dospělou populaci a nejvyšší věkové skupiny

Doctoral Thesis

Thesis Supervisor: RNDr. B. Burcin, Ph.D.

Prague, 2012

Declaration:

I declare that this Thesis is my own work and that I cited all the used sources of information or literature. This Thesis or its substantial part has not been submitted to obtain another or equivalent academic degree.

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

In Prague/V Praze, 6. 4. 2012

Signature/Podpis

Acknowledgments

Before the words of thanks will come I would like to point out that the preparation of this Doctoral Thesis got during the years 2010 and 2011 a financial support from the Grant Agency of the Charles University in Prague (GA UK No. 0136/2010, Methods of mortality analysis focused on the oldest age-groups). I am grateful for this opportunity which enabled me to attend several scientific courses or conferences within the Czech Republic as well as abroad. All these experiences are very important for me and will be also in the future.

At first, I would like to express my gratitude to my supervisor, dr. Boris Burcin for all his comments, inspiration, motivation, ideas and help during the latest years as well as for all the sleepless nights when this Thesis was completed. In the same time I would like to thank also to all my colleagues who ever gave me some advices, support, motivation and words of hope in my hopeless situations. Above all I am grateful to my friends, schoolmates and colleagues Olga Sivková and Petr Mazouch who spent many hours listening and discussing my thoughts, ideas and hypotheses. Anytime they gave me the feedback which I needed. I am glad I experienced the fruitful discussions in Rostock, above all with Trifon Missov, Adam Lenart, Vladimir Canudas-Romo, Grifith Feeney, Nadine Ouellette, Sergi Vidal Torre and of course James Vaupel and many others I forgot to mention here. I would like to thank also to all my great students for their patience, enthusiasm and fresh and inspiring ideas.

Unfortunately, I cannot anymore thank to all my colleagues who were significantly important for my work. At this place I have to mention doc. Felix Koschin, great teacher and person. My regret is that I will never hear his opinion about this work.

My greatest gratitude goes to my great family which I can ever rely on. I was blessed by my unbelievable parents with their infinite patience, tolerance and belief in me. Without them I would never be able to do what I do. I can never express fully how grateful I am also to my beloved husband Vladimír. During the latest months and years he proved many times to be the right person for the moments of crisis, he helped me and gave advices when I needed it and shared the happiness when there was a reason.

Thank you all!

Poděkování

Před samotným poděkováním je třeba zmínit, že tato práce získala pro roky 2010 a 2011 finanční podporu z Grantové agentury Univerzity Karlovy v Praze (GA UK č. 0136/2010, Způsoby zkoumání procesu úmrtnosti se zaměřením na nejvyšší věkové skupiny). Jsem za tuto příležitost vděčná, protože mi umožnila mj. absolvování mnoha odborných kurzů a konferencí nejen v České republice, ale i v zahraničí. Všechny zkušenosti získané během těchto let jsou pro mne velmi důležité a budou i nadále.

Nejprve bych ráda vyjádřila můj vděk především vedoucímu mé dizertační práce, RNDr. Borisi Burcinovi, Ph.D., za všechny jeho poznámky, inspiraci, motivaci, nápady a pomoc během posledních let, stejně jako za všechny probdělé noci během dokončování této práce. Zároveň bych chtěla poděkovat i všem ostatním mým kolegům, kteří mi dávali rady, podporu, motivaci ale i naději, když jsem se cítila úplně beznadějně. Především jsem vděčná mým přátelům, spolužákům a zároveň kolegům, Oli Sivkové a Petru Mazouchovi, kteří strávili hodiny posloucháním a diskutováním mých myšlenek, nápadů a hypotéz. Vždy mi poskytli potřebnou zpětnou vazbu. Ráda jsem i za možnost zažít plodné diskuse v Rostocku s lidmi jako jsou Trifon Missov, Adam Lenart, Vladimir Canudas-Romo, Grifith Feeney, Nadine Ouellette, Sergi Vidal Torre a samozřejmě James Vaupel a spousta dalších, které zde nemůžu vyjmenovat. Děkuji také mým skvělým studentům za jejich trpělivost se mnou, jejich zájem a mnohé svěží a inspirující nápady.

Bohužel již nemůžu poděkovat všem kolegům, kteří významně ovlivnili mou práci. Na tomto místě bych alespoň ráda zmínila doc. Felixe Koschina, skvělého učitele i člověka. Je mi líto, že se nikdy nedozvím, co by si o mé práci myslel.

Mé největší díky směřují samozřejmě k mé skvělé rodině, na kterou jsem se za všech okolností mohla spolehnout. Mám štěstí, že mám tak neuvěřitelné rodiče s nekonečnou trpělivostí, tolerancí a vírou ve mně. Bez nich bych nikdy nemohla dělat to, co dělám. Také asi nikdy nedokážu plně vyjádřit, jak vděčná jsem mému milovanému manželovi Vladimírovi. Během posledních měsíců i let mnohokrát dokázal, že je tím pravým člověkem pro zvládání krizí, pomohl mi a poradil, když jsem to potřebovala a sdílel mou radost, když k ní byl důvod.

Děkuji vám všem!

Selected methods of mortality analysis focused on adults and the oldest age-groups

Abstract

Questions about human life span, longevity and mortality in general are natural to almost everyone. This Doctoral Thesis deals with one central question – whether some limit of human life span or of its improvements exists. It is rather a methodological work, therefore its aim is to introduce not only relevant theories but above all the methods usable in the mortality analysis focused on adults or the oldest-old. At the beginning the most important theories and opinions of scientist dealing with mortality are introduced. In the first half of the analytical part mainly the traditional and basic approaches are included. The theme of life tables is opened by an analysis of its construction in the Czech Republic, together with its possible modifications. As a result the independent macro code for the SAS software is attached in the electronic Appendix. This macro enables to calculate the unknown parameters of selected mortality laws by the method of weighted non-linear least squares and to produce the smoothed and extrapolated values of mortality rates. Using the individual life durations, life tables according to education attainment were constructed (also attached in the electronic Appendix).

In the second half of the work, there are several more sophisticated methods introduced. The first of them is analysis of the rectangularization process which is closely related to the life tables. The process is studied through several defined indicators. Many of them show the stagnation of their values during the latest years. Based on the results, the terms "derectangularization" and "shifting" were distinguished clearly. The process of mortality shifting was analyzed more in depth. Existence of this process is also the most important assumption in the analysis of tempo effects. In this work also the situation when this assumption does not hold was solved. Finally, the last methodological chapter was devoted to the frailty models, a concept related more to cohort data.

Only several methods were incorporated to this Thesis, those using period data. That opens the themes for future research in this field, oriented more on cohort and also more detailed or even individual data. All the methods in the Thesis were also applied to real data. Where possible, data from Eastern as well as Western or Northern European countries were used.

Key-words: mortality analysis, graphical methods, mortality laws, life tables, rectangularization process, tempo effect, frailty models, SAS

Vybrané způsoby zkoumání procesu úmrtnosti se zaměřením na dospělou populaci a nejvyšší věkové skupiny

Abstrakt

Otázky týkající se dlouhověkosti, délky lidského života a úmrtnosti jsou přirozené asi všem. Tato dizertační práce vychází z klíčové otázky – zda existuje nějaký limit lidského života nebo jeho růstu. Práce je zaměřena metodologicky a jejím cílem je nejen představit relevantní teoretické přístupy, ale především metody využitelné v rámci analýzy úmrtnosti zaměřené na dospělou populaci a nejvyšší věkové skupiny. V úvodu jsou zmíněny nejdůležitější teorie a názorové proudy odborníků zabývajících se úmrtností. V první půlce analytické části práce jsou obsaženy spíše základní a tradiční přístupy. Tématika úmrtnostních tabulek je uvedena popisem konstrukce používané v České republice, včetně možných modifikací. Výsledkem je vytvořené samostatně použitelné makro pro statistický software SAS (elektronická příloha), které umožňuje metodou nejmenších vážených nelineárních čtverců počítat odhady hodnot neznámých parametrů vybraných úmrtnostních funkcí i vyrovnané a extrapolované hodnoty měr úmrtnosti. Za využití výpočtu individuálních délek lidského života byly sestaveny úmrtnostní tabulky podle dosažené úrovně vzdělání (elektronická příloha).

Ve druhé půli práce jsou zařazeny pokročilejší metody analýzy. První z nich je analýza procesu rektangularizace, úzce svázaného s úmrtnostními tabulkami. K analýze bylo definováno několik ukazatelů, hodnoty mnoha z nich vykazují v posledních letech spíše stagnaci. Na základě dosažených výsledků byly důsledně odlišeny pojmy „derektangularizace“ a „shifting“ (posun úmrtnosti). Proces posunu úmrtnosti je v další části analyzován samostatně. Existence tohoto jevu je zároveň základním předpokladem v analýze vlivu tzv. časování. Tato práce se však zabývá i situací, kdy tento předpoklad dodržen není. Poslední metodologická kapitola je věnována modelům křehkosti. Jedná se o koncept více spojený s kohortním typem dat.

V práci je obsaženo jen několik vybraných metod, byly voleny ty, které pracují s průřezovými daty. To zároveň otevírá tématiku vhodnou k dalšímu rozpracování v oblasti analýzy úmrtnosti, která by se měla více zaměřovat směrem k využití kohortních dat a také detailnějších nebo dokonce individuálních údajů. Všechny zahrnuté metody byly aplikovány i na reálná data. Kde to bylo možné, bylo do analýzy zahrnuto co nejvíce zemí, nejen z východní části Evropy.

Klíčová slova: analýza úmrtnosti, grafické metody, zákony úmrtnosti, úmrtnostní tabulky, proces rektangularizace, efekt časování, modely křehkosti, SAS

CONTENTS

LIST OF ABBREVIATIONS

HMD	Human Mortality Database (www.mortality.org)
IQR	Interquartile range, one of the basic demographic indicators used traditionally in the analysis of rectangularization process. It is the difference of lower and upper quartile of ages at death in a given population or generation.
LE	Life expectancy
MAD	Mean age at death based on the mortality rates of the 2nd kind (reduced). See Chapter 9.
$\mathbf{MAD_L}$	Mean age at death based on the mortality rates of the 1st kind. See Chapter 9.
$\mathbf{MAD_L^*}$	Tempo-adjusted mean age at death based on the mortality rates of the 1st kind. See Chapter 9.
NPC countries	Group of non-post-communist countries used in the analysis in the 8th Chapter. In this group of countries there are Australia, Austria, Belgium, Canada, Germany, Denmark, Spain, Finland, France (total population), the United Kingdom (total population), Switzerland, Chile, Ireland, Iceland, Israel, Italy, Japan, Luxembourg, the Netherlands, Norway, Portugal, Sweden, and the USA.
PC countries	Group of Eastern or Central European post-communist countries used in the analysis in the 8th Chapter; those are Bulgaria, Belarus, the Czech Republic, Estonia, Hungary, Lithuania, Latvia, Poland, Russia, Slovakia, Slovenia, and Ukraine.
TMR	Total mortality rate. This measure is the corresponding to the generally known total fertility rate. Total mortality rate could be calculated as the sum of the mortality rates of the 2nd kind. See Chapter 9.

LIST OF TABLES

LIST OF FIGURES

... even till death
is not forever ...

Markéta Procházková-Lutková, So Long / Tak dlouho

Chapter 1

Introduction

The issue of senescence and population aging, taken from whichever point of view, is the theme of many actual scientific and (not only) demographic studies. Estimations and forecasts of survival to the oldest age groups are among others also the political need because many developed societies need better planning so as to avoid the bankruptcy of their social, medical, insurance, pension and other public systems (Wachter, 1997). However, an interest on mortality and possibilities of future life span extension are natural to everyone. Biologists, demographers and many other scientists are trying to find any explanation of various individual life durations, causes of differences and factors which possibly influence the human life span. The main reason behind this is the aim to find any ways how human lives could be extended (Gavrilov, Gavrilova, 1991).

In the past, the main reason for population aging was the decline of fertility. In contemporary societies, the process of population aging is mostly determined by the decline of mortality, mainly at older ages (Gavrilova, Gavrilov, 2011). That is why still more attention in demography is devoted to the methods of mortality analysis focused mostly on the higher age groups. Accurate estimations of the future development of mortality at higher ages are then a key factor of accurate forecasts of population aging (Gavrilova, Gavrilov, 2011, p. 3). The subject of interest of this Doctoral Thesis is not only the mortality development but above all the description and study of possible methods of mortality analysis of adults and especially the oldest-old people.

Already at the end of the 20th century one half of dying women and one third of dying men died after their 80th birthday (Kannisto *et al.*, 1994). In demographically developed countries that are the "oldest-old" (see the definition below) whose share in the population is the fastest growing during recent years and the increase is likely to continue, and whose mortality rates are significantly changing. Even among demographers after the mid of the 20th century, it was a common belief that mortality rates are already nearing some minimal limit (as mentioned e.g. Gavrilova, Gavrilov, 2011), and that no significant improvements of those rates are practically possible. Research within this issue was complicated with relation to the lack of reliable data of sufficient quality. Only on the basis of more reliable data on mortality of the oldest people it has been shown that even in developed countries with general low mortality level it had been achieved a significant decrease in the intensity of mortality, also at the highest ages (*ibid*.). The dimension of this positive trend was growing during the 20th century; Kannisto (1996) estimated the average decrease in mortality rates in the European countries (except for the Eastern Europe) from the 1960s to the 1990s for women to be ca 1–2 % per year and for men about 0.5–1.5 % per year. As a result of these developmental trends, there are more people surviving to higher ages, thus the database suitable for the study of this issue is expanding and of course there is also an increased interest and opportunities of demographers to analyze this theme. On the other hand, for the same mentioned reason, also the need for such type of demographic analyses in practice increases.

Despite the aforementioned improvements of available data, data on mortality of people in the highest age groups still could be classified as less reliable and rare. In such cases, when it is impossible to carry out a comprehensive analysis based solely on published data, it is necessary to use some simplifying assumptions, generally to construct specific models. It is exactly that part of demography covered partially in this work what is often associated with demographic models, formal approaches, and general expressions, which are often validated only ex post using the empirical data. This approach thus partly falls into the field of so-called formal demography, working primarily with formal expressions and formal treatment of demographic relations and models. In case that some of the necessary data become accessible, it is possible to fill the data to the formally expressed estimated relationships and so to verify the validity of those relations.

Formal demography is an area of demography that is rapidly developing and has its representatives in various parts of the world and within various fields of demography itself. This is despite the greater availability of data these days, or even it could be said that this is so just because of the increased availability of data. Better and more detailed data helped to reveal many parts of reality, which were not studied by demographers or even which could not be studied before. An example of such tendencies may be the effort to analyze the shifts of the survival curve (process of shifting of the survival curve), theory of population heterogeneity, or considering the so-called timing effect in mortality analysis and many other areas, some of which are also included in this Thesis.

1.1 The aim of the work

Demography is mainly a quantitative discipline with important interdisciplinary character. Some of its methods and concepts have not been questioned until recently (Luy, 2010a). In accordance to the above mentioned trends in the world's demography, the main goal of this Thesis is to introduce, analyze or even deepen several selected methods of mortality analysis, some of them are often taken as the basic ones in demography, and to show their strengths, weaknesses and possibilities for current demographic research. In other words, within this work several questions will be set and the aim is to try to find some possible answers and solutions to these questions rather than to solve them in one directive style. In this way several basic and simple concepts of demography will be analyzed more in detail, where possible some aspects will be derived or verified on available data. The aim of the Thesis is to show that also the simplest issues could be an interesting source of inspiration for a further research in mortality analysis and sometimes could reveal new or maybe also a bit controversial themes, introduced mainly in the second half of this Thesis.

The second part of the work, as was said already, will leave the basic concepts and build upon them with the aim to present and evaluate some more sophisticated methods. Particular chapters of the text may seem to be almost independent but in fact they are logically tied expressing various views to the same issue – the human mortality. Generally speaking, various methods introduced or used in this Thesis represent various answers to only one initial question about the mortality development in the past, these days, and also theoretically in the future. All the presented methods would lead to more detailed knowledge of the background mechanisms predetermining the future developmental trends.

In accordance to the facts mentioned in the introductory part of this chapter the main objectives of this work can be defined as follows. This Doctoral Thesis attempts to provide an overview of several selected methods usable for the analysis of changes in mortality at adult ages with a particular focus on people at the oldest age groups (i.e. the "oldest-old" – see definition of age groups considered later in this chapter), including their application in demographic practice. The goal is not to provide an exhaustive summary of all possible and existing analytical methods, the attention will be devoted particularly to several issues and methods that are currently the most frequently discussed in demography, are actually studied or are promising for the future usage, and sometimes are even slightly controversial in the field of demography. Moreover, the included methods use period data for the analysis. Some of the methods may seem traditional and basic ones, others are rather new or developing. It is supposed that the reader of this work has knowledge of at least the basic terms and methods commonly used in the mortality analysis. This text cannot be understood (and even that was not its intent) as a sort of textbook of basic analytical approaches to mortality. From the text below it is rather possible to build upon these basic approaches, enrich them and possibly to put them into the broader framework of adequate methods. It only shows the extent of analytical tools that are now usable for demographers. These tools are still becoming more enriched and supplemented; therefore, to give a comprehensive overview of all possible methods of mortality analysis with all their modifications and derived approaches is not even technically possible at all.

Methods introduced and mentioned within this work involve both the analytical and graphical approaches to the analysis. The introductions of both those groups of methods are then sub-objectives of this Thesis, and together they lead to the fulfillment of the main objective mentioned above.

Graphical methods (at least the methods included in this Thesis) allow effective comparison of the longitudinal (cohort) and transversal (period, cross-sectional) point of view to the studied issues by surface (so-called mortality surfaces) or three-dimensional image of the studied phenomenon. Both these ways of illustration are easily handled by available software; nevertheless, those are not yet a standard part of scientific outcomes (especially in the Czech demographic practice). So **the 1st partial goal** of this work is to present briefly the advantages of these methods and the possibility of their integration into routine work of demographers.

Utilization of these essentially simple but illustrative approaches may enable to get more into the very structure of the data during the initial phase of the analysis, to reveal the most important relationships in the data at a first glance, and thus to easily select adequate analytical approaches for further data processing. Of course, it could be said that only the introduction to the existing graphical methods of mortality analysis could be sufficient for an independent work or Thesis. Therefore, for this work it is not the aim to introduce the graphical methods fully. In the chapter devoted to graphical methods only those methods are selected which are then used within the rest of the work so as the reader can easily find how the particular types of graphs were constructed and what are their most important advantages.

Analytical procedures represent the main part of this Thesis and they include a wide variety of methods usable for analysis of trends and changes in mortality rates of adults and the oldest people. Particular methods will be presented separately but with a logical connection. The reader will be led from the simplest issues to the more complicated ones which are, however, based on the former mentioned ones.

The 2nd partial goal of this work is to analyze the current practice of official life tables' construction mainly in the Czech Republic, which uses the traditional Gompertz-Makeham method of mortality smoothing. Although the life table construction could be taken as a simple and clear task in demography, there it is quite a lot of space devoted to this issue in the text of this Thesis. The reason for this is the fact that the life tables and the development of approaches to their construction cannot be considered as completely finished. Even in this area it is still possible to discover new and less traditional approaches. One of them is the possibility of choice from more various smoothing and extrapolation functions or of some methods usable for the estimation of their unknown parameters. Part of this Thesis is focused on a deeper analysis of the construction procedure of life tables with a special focus on contemporary practice in the Czech Republic. It describes the procedure used for parameter estimation of the Gompertz-Makeham function in the Czech practice and shows the sensitivity of this procedure to the initial assumptions.

Next part of the Thesis will focus on alternative ways of smoothing and extrapolation of the mortality curves, which might better reflect the current situation of mortality in developed countries. There have been several dozens of these approaches developed, but not all of them are universally applicable, in addition they give significantly different results at the highest ages. Except of the necessary choice of a proper method a demographer while constructing the life table has to select and use proper method of parameter estimation. There are various possibilities; however, usually it holds that the more accurate method the more complicated its application is. Therefore, **the 3rd partial goal** of the Thesis is then the introduction of one of the most accurate methods of parameter estimation, the non-linear weighted least squares method. Its application could be quite complicated and difficult in practice. Moreover, in demography it is often necessary to repeat the calculation many times – for more years (or more populations in general), or for more different methods of smoothing. In connection with the need to automate and accelerate this calculation more, a program code for the SAS statistical software was developed and attached in an electronic form to this Thesis as an Appendix. An alternative could be the new independent statistical software DeRaS[1] (Burcin *et al.*, 2011b; Burcin, Hulíková Tesárková, 2011) which enables not only to choose from several mortality laws (or smoothing functions) but also to produce final life tables constructed on the basis of the selected method of smoothing and extrapolation and the application of moving averages to lower ages.

The next part of the Thesis then is devoted to the introduction of the so-called process of rectangularization of the survival curve. In the Czech Republic, there this process could be studied on the basis of current life tables (it is also possible to use alternatively constructed life tables – see previous paragraph) or tables downloadable from the international databases (e.g. Human Mortality Database[2]). Data from the Human Mortality Database will be used throughout the whole Thesis where some international comparison will be done. The first results comparing the Czech Republic and other selected countries were already published (Burcin *et al*., 2009; Šídlo, Tesárková, 2009). These results were published within the preparation of this Doctoral Thesis and so they became also the basis of the chapter devoted to this topic. Part of the chapter also deals with the basic findings related to the so-called shifting process (the shift of the survival curve or other table functions). Sources cited in that part indicate that particular authors do not understand that term consistently yet. Therefore, in theoretical terms, some possible approaches to this process are shown and formal relationships are derived, which illustrate the possible reaction of the table functions to the change of intensity of mortality. The derived relations were also verified using the empirical data. In accordance to that, **the 4th partial goal** of the Thesis is to present the issue of the rectangularization process and its consequences, and mainly to deal with the process of mortality shifting in the theoretical as well as practical point of view.

Particularly the analysis of the rectangularization process (or process of mortality compression) is one of the most common approaches to investigation of the changes in mortality conditions in the world. It can also contribute to the search for the answer to one of the oldest of human questions – whether there any limits of human life span exist and whether the human population is approaching some limit (Wilmoth, 1997). The 3rd chapter of the Thesis is devoted to the definition and description of selected basic theories related to this issue.

Study of the impact of tempo (timing) effect to the mortality analysis, could be considered almost as a controversial issue in demography. The assessment of such theories may lead to consideration of the usage of other, alternative, indicators than are those rather traditional and commonly used ones (analysis of fertility underwent also a similar change, or enrichment, of approaches). An indivisible part of the Thesis, devoted to the tempo effects, is also an introduction to basic assumptions on which the adjustment of standard indicators for the so-called tempo is based and its verification on real data. Using historical data accessible for the analysis of the tempo effect it turned out that one of the basic assumptions of this approach (assuming a parallel shift of the survival curve over time – the shifting process – without changing the shape of this curve or of the hazard curve), especially in the past cannot be regarded as fulfilled. This fact could significantly influence the results. Therefore an alternative approach that allows including of the assumption of changing variability of mortality with age is

[1] http://deras.natur.cuni.cz/
[2] http://www.mortality.org

introduced too. This alternative approach, originally developed for the process of fertility, was slightly modified within this Thesis to be suitable for the mortality analysis. **The 5th partial goal** of the Thesis is thus the presentation of this concept and evaluation of its suitability for available data related to the process of mortality.

Yet not fully enforced or known approach is the analysis of population heterogeneity, although its roots can be traced back several decades. The relatively recently increased possibility of verifying of its basics motivated demographers for further elaboration of this approach in order to uncover some internal regularities of the population and to contribute to the detailed results of previously performed analyses. **The last, 6th, partial goal** of this Thesis is thus to offer a short introduction to this concept and a short model illustration of its usage.

It may seem that from the geographical point of view the Thesis is focused mainly on the Czech Republic. But the true is that in some cases data of the Czech Republic are used only for the illustration of a particular method which could be then applied to any other data set. Where more countries are analyzed then usually more Eastern European countries are involved into the analysis so as the comparison with Western or Northern Europe could be done. The greater focus on the Eastern part of Europe or post-communist countries resulted partially from the fact that those countries are often missed out in the original articles or studies introducing the modern demographic methods and partially from the logically greater interest of the author.

1.2 Division and definition of age groups used for the adult population

Groups of people at the highest ages are defined in different ways in literature and delimited on the basis of differently age-defined borders. Some of the most general and frequently used terms in practice are the terms “seniors” or “persons at a post-productive age”, usually both these terms refer to persons aged 65 and over (Burcin *et al*., 2003). The “post-productive age” follows the division of the population to the so-called economic generations: pre-productive generation, productive and post-productive (e.g. Roubíček, 1996 or Pavlík *et al.*, 1986). In this case, a lower age limit of the post-productive age can be defined depending on the retirement age in the studied population. That is the reason for its possibly different definition in the past in comparison with the present. Today, the most frequent lower border of the post-productive generation is the age of 65 for both sexes. For the purpose of this work thus the term “post-productive generation” defines the same age group as “seniors”, i.e. persons aged 65 years and older.

In older demographic works, a group of people over 65 is often analyzed as a whole, because the number of people, who survived to this age, although it was not negligible, was not large enough to allow more detailed division of this group. In recent years the number of elderly people increased enough (Kannisto *et al.*, 1994) that in demography a more detailed division of people over 65 years of age started to be applied. The group of the oldest people (aged 85 and more years) is denoted as “the oldest-old” (e.g. Campion, 1994), people aged from 65 to 84 are often divided into so-called “young-old” and “old-old” (e.g. Atchley, 2004, cit. according to Kafková, 2006 or Chou, Chi, 2002, and others). Other denominations (originally in Czech) states e.g. Hazafyová (2009) or Mühlpachr (2004, as cited in Jůzová, 2006), it is the division into "young seniors" (65 to 74 years), "old seniors" (75 to 84 years) and "very old seniors" (85 and over).

In the literature devoted to this theme, there are sometimes people aged 100 years and older (referred to as "centenarians") analyzed separately and, possibly, also people aged 110 years and older (the so-called "supercentenarians").

1.3 Current state of knowledge related to the theme of the work

Because this Doctoral Thesis introduces almost independent issues, there will be a more detailed summary of relevant published literature presented within the particular chapters, which deal with the individual topics. As well as, the current state of knowledge on each topic will be more discussed in all the relevant chapters of this Thesis. In this chapter, there are the most important roots of the themes mentioned and related works which have been presented or published in some form (scientific article, conference paper, teaching) by the author while working on the Thesis (i.e. during years 2008–2012).

The issue of human longevity and mortality at the highest ages is the subject of research for an increasing number of demographers all around the world. For centuries, the traditional tool of mortality analysis in demography has been the life tables, a model expression of the mortality process in the studied population. The base for a construction of such tables is also, among other calculations, the selection and application of one of the methods of smoothing and extrapolation of mortality rates or the probabilities of dying, that means the estimate of the force (intensity) of mortality. The most traditional approach used today for example in the Czech statistical practice is the usage of the moving averages and Gompertz-Makeham function (Czech Statistical Office, 2009b). In general, this method was fully satisfactory at least until the middle of the 20th century. Then mortality rates began to improve rapidly and the trend of lengthening of the life expectancy became more important. At the time of the increasing survival of people to older ages this approach started to be criticized because of overestimations of mortality at the highest ages. It was pointed out also in the Czech demographic literature by e.g. Koschin *et al.*, (1998) or Koschin (1999). Some possibility of modification of the traditionally used calculations within the life tables was already shown by Pecka (1989). Based on changes of the mortality development (above all) at the highest ages, many other relationships and laws were formulated that can be used for the smoothing and extrapolation of the mortality rates at these age groups. These functions include for example in the Czech journal *Demografie* introduced the so-called modified Gompertz-Makeham function (Koschin, 1999); from foreign publications the cubic model (Pastor, 2007), as well as models bearing names of their authors, for example Heligman-Pollard (Boleslawski, Tabeau, 2001), Coale-Kisker (Boleslawski, Tabeau, 2001; Coelho *et al.*, 2007), Denuit-Goderniaux (Coelho *et al.*, 2007), Thatcher (Thatcher, 1999) or model of Kannisto (Thatcher *et al.*, 1998; Coelho *et al.*, 2007), used also within the Human Mortality Database project. It is the latter one model which is very popular among demographers because of its clarity, ease of calculation and quality results, which provides (Roli, 2008;

Yi, Vaupel, 2003; Burcin *et al.*, 2010). In terms of suitability and accuracy of expression of various differently constructed models there is no clear consensus among demographers. Most of these methods were not described in the Czech demographic literature until recently, or applied to the Czech data. In the context of the preparation of this Thesis, the issue of smoothing and extrapolation of the mortality rates and the probability of dying for the life table construction became the theme of the article in the Czech journal *Demografie* (Burcin *et al.*, 2010).

The life table construction is closely related to the analysis of the process of rectangularization of the survival curve. The survival curve is one of the outputs of a standard life table. The rectangularization process can be described as an approaching of the shape of this curve to a rectangle, which is caused by an increasing proportion of population surviving to higher ages. Their deaths are then more concentrated into a shorter period of age (because of that the process is also known as a compression of mortality). This process was studied by biologists at first, in demography it was implemented most visibly by James Fries (1980). To analyze this process many indicators, indices and benchmarks can be used. The partial list of them was prepared while working on this Thesis, and it was published together with particular results for the Czech Republic (in comparison with other mostly European countries). The results were introduced in several presentations in the Czech Republic and abroad (Burcin *et al.*, 2009; Šídlo, Tesárková, 2009).

An integral part of the mortality analysis is the evaluation of the level of mortality depending on age and its changes over time. Demographers have dealt this issue basically since the beginning of demography as a separate scientific discipline. Today’s view focuses for example on the effect of timing (Bongaarts, Feeney, 1998; 2002 or 2006) or population heterogeneity, an important representative of this direction is e.g. J. Vaupel (Vaupel, Yashin, 1983; Vaupel *et al.*, 1979; Wienke, 2011). Population heterogeneity can be simply thought of as taking into account the diversity of the population in demographic analysis. It is based on the logical fact that people vary in terms of probability of dying. If we accept the assumption that the heterogeneous population can be divided into a number of more homogeneous ones, the overall mortality pattern is at each age basically equal to a weighted average of the intensities of individual sub-populations. The overall force of mortality can be considerably different from the intensity of each sub-population due to changes in population structure (i.e. change of the proportion of population with higher or lower intensity). Practical examples of heterogeneity effect provide Vaupel and Yashin (1985). To this concept the article in the Czech journal *Demografie* was devoted (Hulíková Tesárková, 2012).

Study of mortality in terms of transversal (cross-sectional, period) or longitudinal (cohort) point of view is not new in foreign and in the Czech literature either. However, it should be noted for example that the cohort life tables for the Czech Republic have not been published yet. It is possible to illustrate the cohort development easily thanks to special methods of graphical presentation that are available through the usage of modern software, such as the SAS software or R. Those graphical methods could be for example the mortality surfaces illustrating the probability of dying or other functions according to age over time using colors for the differentiation of various levels of its values. These procedures can be used not only for the study of current intensity of mortality, but also for the illustration of its changes in time, for the analysis of differences between states, assessing the impact of selected causes of death, etc. (Barbi *et al.*, 2004). Due to its simplicity and clarity, these methods are very popular. An alternative to the distinguishing among various levels of intensity by colors could be the usage of the 3-dimensional graphs (they again could be created easily using the appropriate software). These types of illustration of the studied variables and their development will be used throughout the Thesis as well as in particular publications related to it (*e.g.* Šídlo, Tesárková, 2009; Hulíková Tesárková, 2011).

Nobody grows old merely by living a number of years.
We grow old by deserting our ideals.
Years may wrinkle the skin,
but to give up enthusiasm wrinkles the soul.

Samuel Ullman

Chapter 2

Selected graphical methods usable in the mortality analysis

2.1 Graphical methods in demographic analysis

For the first basic analysis of data some relevant graphical methods could be used. The aim of the graphical analysis is not a detailed quantitative description of the mortality process of the population. However, suitable graphical methods could enable to describe and to study the overall trends of the studied process and to consider it from more points of view. There are many possibilities of graphical analysis of the data, most of them could be prepared in a basic and widely available software. In this chapter just those methods which enable to see not only the development in time or the dependence on age but both those dimensions will be presented. Figures which can show the development in time and also across age are corresponding with the Lexis diagram. This type of graphs could be prepared also in the commonly available software, for the purpose of this chapter the possibilities of the statistical software SAS and R will be shown illustratively.

Because many of the analyses in the Thesis were done in the SAS software, this software was also selected for the presentation of some results. The second selected software used here is the software R, downloadable free of charge on the Internet. The reason of this selection was its accessibility and also it’s growing popularity for demographic research.

The aim of the chapter is to show the possibilities of these methods and a brief description of their usage and preparation. All the graphs in this part are shown only for the illustration and detailed description of the results will be presented in particular parts of this Thesis dealing with the particular themes. Later in the work the graphical methods introduced here will be used without its technical description and explanation.

2.2 Mortality surfaces

Mortality surfaces are example of simple graphical methods which enable to study both the dimensions of data – the development in time and in dependence on age. The main idea is the same as for the standard Lexis diagram traditionally used in the demographic analysis. In a mortality surface the horizontal axis represents calendar years and vertical axis shows the age. In the area of the graph, there is the intensity of the studied process – for example values of the mortality rates. Different values of intensities are distinguished by different colors. Theoretically values for triangles of the Lexis diagram could be shown in the graph or values for all three sets of the Lexis diagram. In our examples the 3rd class of events is used (squares in the Lexis diagram).

Mortality surfaces could be easily prepared in many types of software; within this chapter the software R was used (R Development Core Team, 2009). In R a user has to write the programming code. For preparation of the mortality surface the most important part of the code has to contain the following rows[3]:

```
Age<- c(0:100)
```

which defines the values of age used in the graph. In this case we will use values from 0 to 100 years. These values are stored in an arrow named “Age”.

```
Year<- CZE[1,]
```

The second row defines the range of values of calendar years shown in the graph. In this case it is defined as the first row (“[1,]”) of the table called “CZE”. This row contains the values of the calendar years for which data are available. The values are stored again in an arrow named “Year”.

```
levels=c(0,.0001,.0005,.001,.005,.01,.05,.25,1)
```

In the third row values which will be distinguished by different colors are set. Values above one appeared only in a few rare cases at the highest ages. These cases are not shown in the graphs. The distinguished values could be changed easily according to the needs of the data.

```
filled.contour(Year,Age,t(CZE),levels=levels,ylab="Age",
xlab="Year")
```

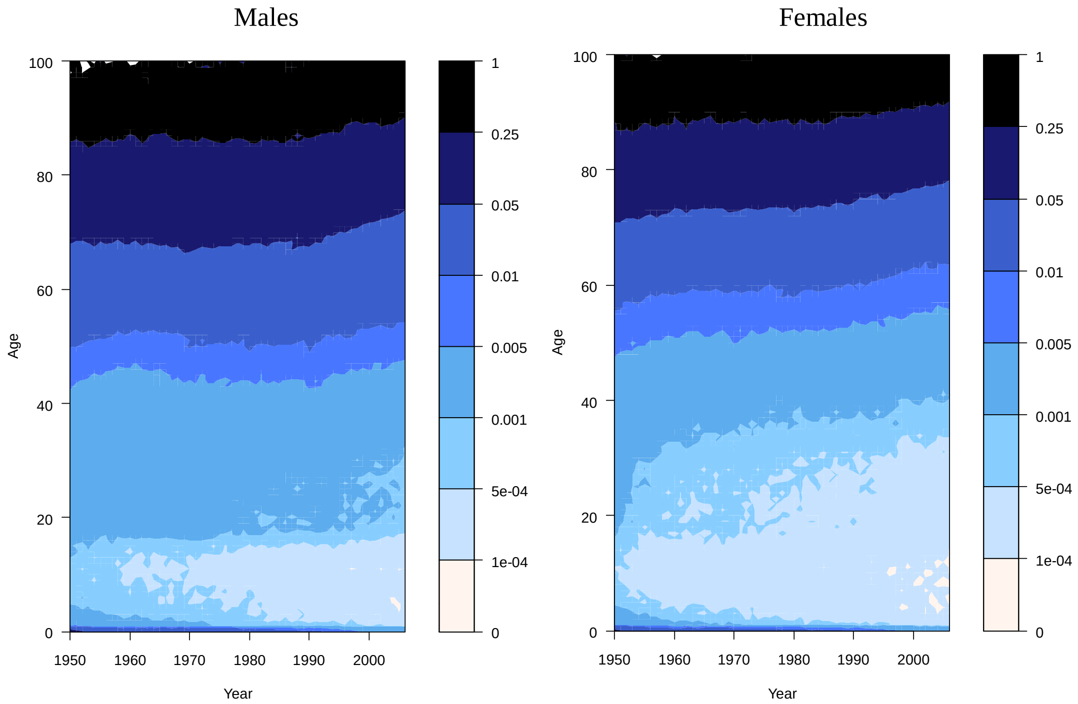
The last row then constructs the mortality surface (“filled.contour”). Within the brackets then both the axes are defined and the values showed in the graph (taken from the table “CZE”), levels and labels of the axes are set. To the code also many other specifics could be added (colors, labels, etc.) and the final graph could be more specified.

Results in this chapter are presented only for illustration and that is why they will not be described or analyzed in more detail. For presentation of the graphical methods values of the

[3] Parts of the code introduced here are inspired by Vladimir Canudas-Romo and his R-codes used within the course *Joint IUSSP-MPIDR Summer School “Frontiers of Formal Demography”, Max Planck Institute for Demographic Research, Rostock (Germany),* June, 2009.

mortality rates are shown and for comparison data for the Czech Republic and Slovakia were used in the Figures 1 and 2.

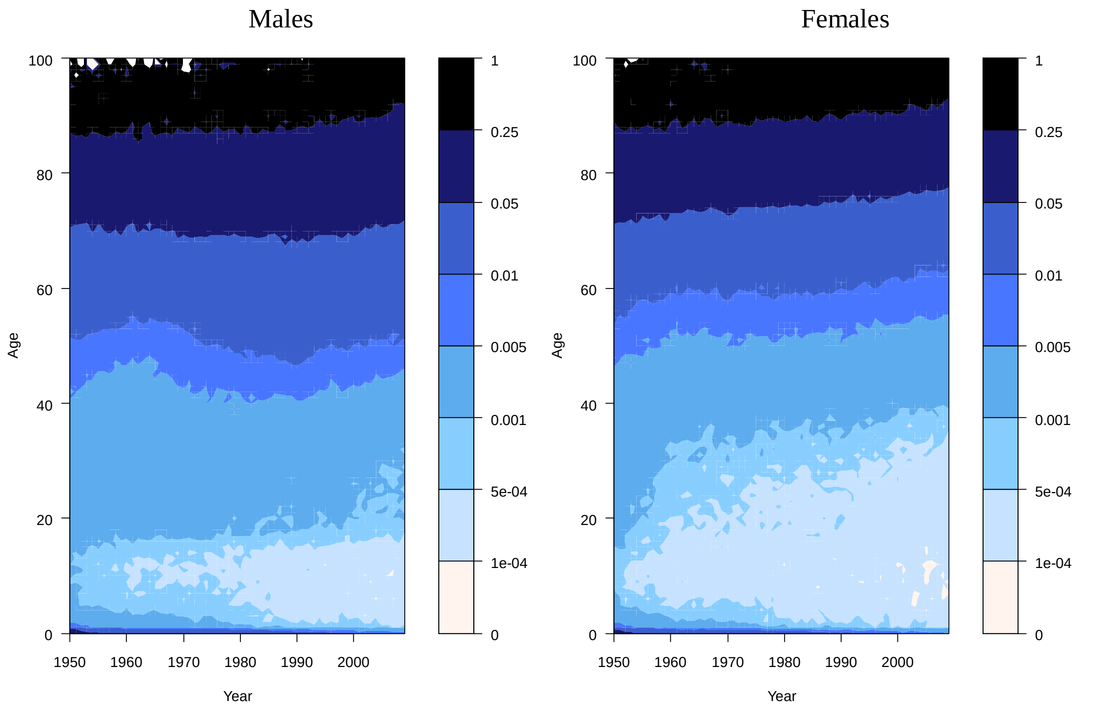
Figure 1: Mortality surfaces, age-specific mortality rates, Czech Republic, by sex, 1950–2006



Note: Prepared in R software

Source of data: Human Mortality Database (2010)

Figure 2: Mortality surfaces, age-specific mortality rates, Slovakia, by sex, 1950–2006



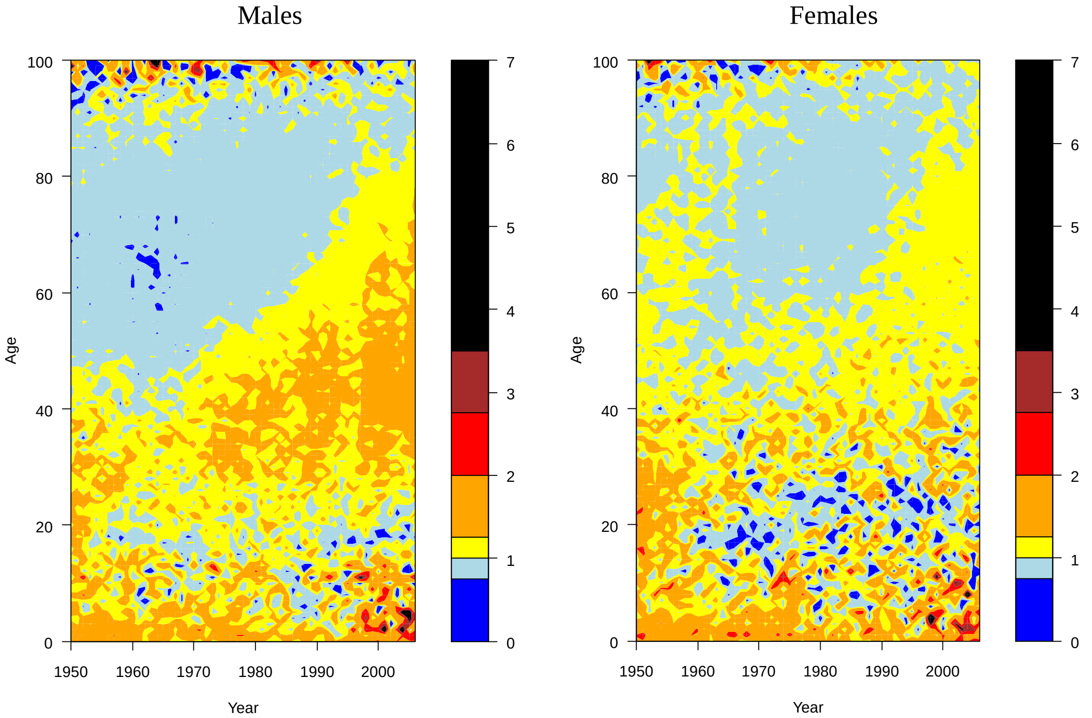
Note: Prepared in R software

Source of data: Human Mortality Database (2010)

One of the advantages of mortality surfaces is the possibility of comparison of the conditions or intensities for two different populations (males vs. females, two countries, etc.). It enables not only the comparison of certain values measured in one moment in time, but also its development in time. For illustration the age-specific mortality rates for the Czech Republic are compared with the development in Slovakia.

For the Figure 3, values for the Slovak Republic were simply divided by the values of the Czech Republic. Then blue colors in the graph signifies lower values for the Slovak Republic than were those for the Czech Republic, the other colors show the advantage of the Czech Republic.

Figure 3: Comparison of the mortality development in Slovakia and the Czech Republic, by sex, 1950–2006



Note: Prepared in R software; values of the age-specific mortality rates for Slovakia are divided by values of the Czech Republic

Source of data: Author's calculation based on Human Mortality Database (2010)

2.3 3-dimensional graphs

Mortality surfaces represented that type of graphs which are 3-dimensional in fact but the third dimension is expressed by different colors. When more detail should be visible then the 3-dimensional graphs could be the solution. Also this type of graphs could be prepared in more types of software but within this chapter the statistical software SAS (SAS Institute Inc., 2009), version 9.2, will be used. The 3-dimensional graphs prepared in SAS software have one very important advantage. They could be produced in the html-format and so presented easily. Moreover the user can rotate the graph or zoom to selected parts of it without the need of programming another part of the code. Thanks to that many details or specifics in the development could be revealed and then studied. In the same time, also the main trends remain visible. The flexible non-static way of presentation of these graphs is very attractive for the audience and it is also very educative when used in the class.

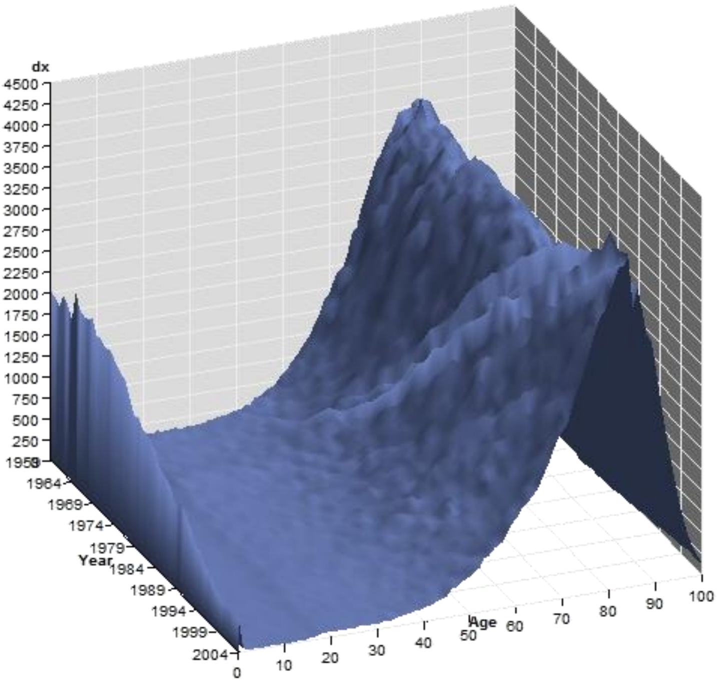
Various angles of rotation enable the presentation and visibility of many specifics features and aspects of the reality, for example:

- effects of specific cohorts (e.g. in the Czech Republic the empirical intensity of mortality is different for people born before, during, and after the World War I);
- shift of the modal age at death to higher ages;
- increasing concentration of deaths around the modal age at death;
- decrease of the infant mortality;
- the main differences among two or more populations, etc.

Also in the SAS software the preparation of the 3-dimensional graphs is not too complicated, a user only has to write the code properly. For the construction there is a procedure named "G3D". The used data set has to be specified after the name of the procedure – in our example it is the file "death_rates_CZE". After the command PLOT all the three axes have to be defined, in our example age and year will be the horizontal axes and values of "dx" (the distribution of table deaths, function d_x in the traditional life tables) will be on the vertical axis. Then the procedure is ended traditionally ("run;" and "quit;").

```
PROC G3D DATA = death_rates_CZE;
PLOT age * year = dx;
RUN; QUIT;
```

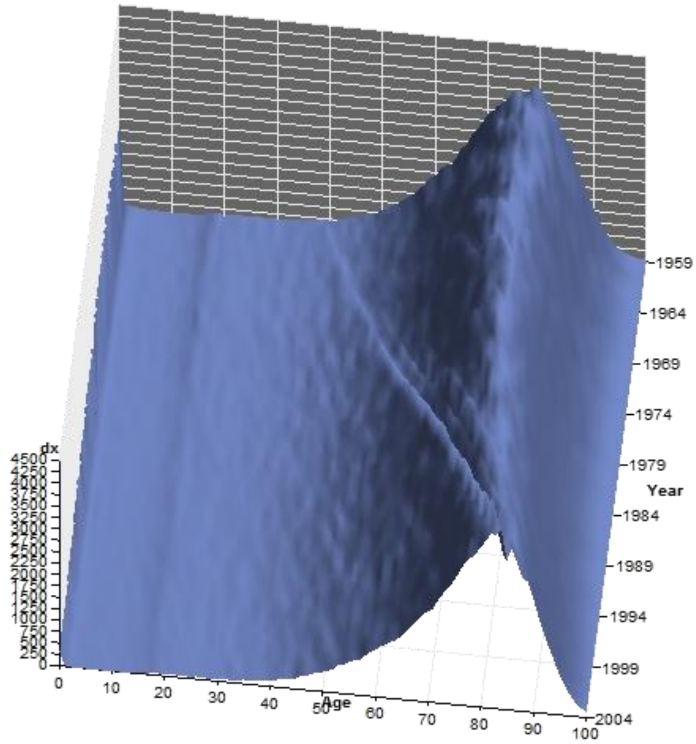
Figure 4: Distribution of table deaths, Czech Republic, both sexes, 1950–2006



Note: Prepared in SAS 9.2 software

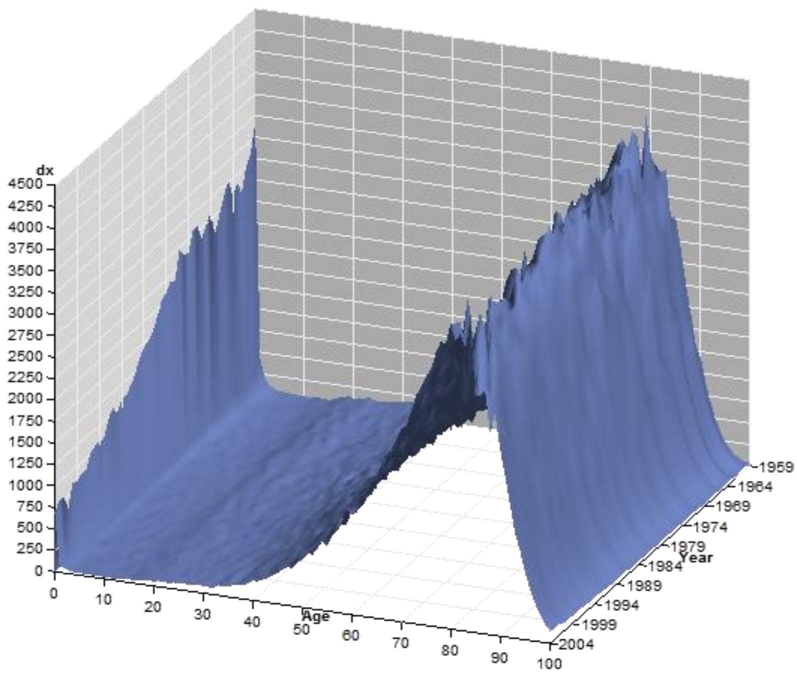
Source of data: Human Mortality Database (2010)

Figure 5: Distribution of table deaths, Czech Republic, both sexes, 1950–2006, different rotation



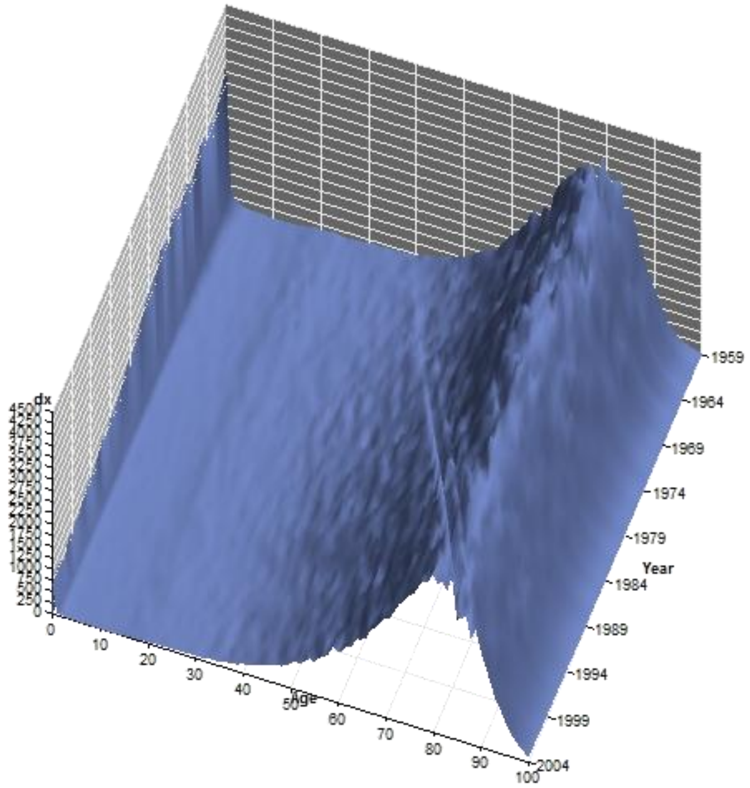
Note: Prepared in SAS 9.2 software
Source of data: Human Mortality Database (2010)

Figure 6: Distribution of table deaths, Slovakia, both sexes, 1950–2006



Note: Prepared in SAS 9.2 software
Source of data: Human Mortality Database (2010)

Figure 7: Distribution of table deaths, Slovakia, both sexes, 1950–2006, different rotation



Note: Prepared in SAS 9.2 software
Source of data: Human Mortality Database (2010)

Of course, many other specifics could be added to the code but the described and shown rows represent the very basic way how this type of graphs could be produced. Moreover, when a graph is already produced its colors, features of the axes, or the overall design could be changed easily.

2.4 Summary

Within this chapter only two of many other possibilities of graphical methods for mortality analysis were presented. These methods were selected on purpose because both of them show three the most important dimensions in demography – age, year and cohorts. Thanks to that it is possible to study the development of mortality in time and according to age without any problems.

The first presented type of graphs, the mortality surfaces, was prepared in the R software but it is possible to use also other software, including SAS. This type of graphs is practical for printed materials because it is not based on the need of any rotation. Another advantage is the possibility of direct comparison of two different populations.

The 3-dimensional graphs presented here (Figures 4–7) were prepared in the SAS software but again it is possible to construct them in other types of software, including R and the latest versions of Microsoft Excel (with significant limitations in comparison to SAS). The main advantage of its construction in SAS it the possibility of many modifications of its design by the user and the attractiveness for presentation.

Both the introduced types of graphs will be used within the whole text of this Doctoral Thesis and were also implemented to most of the publications and presentations connected with this Thesis.

EMILIA: A human shouldn't, shouldn't, shouldn't live so long!
VÍTEK: Why?
EMILIA: A man cannot stand it. To one hundred,
to one hundred thirty years it could stand, but then ...
then he knows ... knows that ... And then his soul dies.
VÍTEK: What does he know?
EMILIA: Oh God, there are no words for that.
And then he cannot believe in anything. In anything.
And that's the boredom.

Karel Čapek, The Makropulos Affair / Věc Makropulos

Chapter 3

Searching for the limits of the length of human life

3.1 Is there a limit of human life?

Rapid improvement of the mortality over the past decades is undoubtedly one of the greatest human achievements (Wilmoth, 1997). This rapid development and important changes that have occurred in the values of all commonly used characteristics of mortality have led many scientists to a question, where the limits of this improvement are. Thus also the question arose what are the limits of how far the length of human life may grow. It is not only a question of maximum attainable age which people are able to live to, but also a question of the quantitative development of the oldest age groups of people with all its consequences for the whole society.

Fries hypothesized that the number of the oldest people in the future will not increase significantly (Fries, 1980). However, as showed for example Kannisto *et al.* (1994), numbers of the oldest people are still increasing. He showed on the basis of reliable data from 27 European countries included in the Kannisto-Thatcher Database on Old Age Mortality[4] that the number of people aged 100 or more years in these countries was lower than 9 000 people in 1980, while 10 years later the number has approached 20 000. Number of people aged 80 years or more then more than tripled between 1950 and 1990 (Kannisto *et al.*, 1994). The oldest groups have become the fastest growing groups in the considered populations. The increase of the number of elderly people is not only the consequence of the decline of mortality rates at the highest ages, but also at lower ages, what enables to bigger proportion of people to survive to the old age. Of course, there must be mentioned the considerable influence of the original size of the generations (Thatcher, 1981, as cited in Kannisto *et al.*, 1994).

Not only demographers, but also biologists, actuaries, and many others are trying to estimate the achievable limits of age for the human population. Despite the considerable importance of a potential limit of life expectancy and its practical implications, scientists have not yet reached

[4] http://www.demogr.mpg.de/databases/ktdb/

the conclusion about its value or even about its existence. Opinions about this question could be summarized into three groups. The first consists of "traditionalists" (e.g. Olshansky, Carnes, Cassel, Fries), who assume that such a limit of life expectancy for both sexes cannot be significantly higher than approximately 85 years. The so-called "visionary" view (as it was called by Manton *et al.*, 1991), or the optimistic view, reflects the assumption that limits of life expectancy will grow even in the future due to ongoing biomedical research. Proponents of this view (e.g. de Grey, Strehler, Rosenberg, Walford) assume a possible increase in life expectancy to values even higher than 100–125 years. The third group is by Manton *et al.* (1991) referred to as "empiricists" (e.g. Vaupel, Oeppen, Wilmoth). They assume that populations are not yet close to the limit of life expectancy as mortality continues to decline and still new successes in the field of medicine are achieved even in relation to the chronic diseases and constraints that play an important role at the end of life. According to this view, it would be possible to achieve values of life expectancy approaching 100 years during the 21st century at the current pace of decline in mortality (*ibid*.).

3.2 Mortality at advanced ages – theoretical background

There exist many theories about the mortality pattern at the highest ages. All of them are closely related to the assumption of the existence of some limit of life span. Vaupel (1997) briefly introduced some of them. He mentioned for example the theory of Buffon, who supposed that "each species has a characteristic maximum life span and that this life span is six or seven times the duration of the period of growth" (Vaupel, 1997, p. 18). Gavrilov and Gavrilova (1991) upon this Buffon's assumption estimated, that for humans the biological life span would be 90–100 years. Also Fries (1980) supposed a fixed limit of life span, as was already written above, and he is probably the most important representative of this ideological direction. Specific theories are based upon the assumption that mortality should grow rapidly at post-reproductive ages where humans have already no evolutionary role (Curtsinger, 1995). Many of these theories are based, above all, on assumptions and ideas rather than on empirical data (Vaupel, 1997).

From empirical data, until recently, the mortality pattern could be studied up to age ca 80 years where there was enough reliable data. During the latest decades more data are available for example in international databases. These empirical data suggest that mortality does not grow exponentially with age at the highest ages (Gavrilov, Gavrilova, 1991; Vaupel, 1997) and it was estimated that mortality probably reaches its maximum around the age of 110. Mortality deceleration was confirmed also in many experiments with other species (Vaupel, 1997). That is the reason why many analytical functions that level off at higher ages were formulated for mortality modeling.

For many demographers and biologists there remains the question what is the reason for the mortality pattern at higher ages. One of the most important ideological trends is the study of the impact of population heterogeneity. It is supposed that the composition of population could lead formally to the deceleration at the end of the mortality curve (Vaupel *et al.*, 1979; Vaupel, Yashin, 1985). This concept will be described in more detail in Chapter 10.

Gavrilov and Gavrilova (1991) mention three sources of variability in the individual life spans: the genetic heterogeneity hypothesis, the stochastic hypothesis and variations in the external environment. Many attempts to verify the influence of heritability (made e.g. by Beeton and Pearson at the beginning of the 20th century, by Pearl some 30 years later or by Philippe during the 1970s) showed nearly the same – that there is almost no heritability of life span in human and also in other species which could be easily interpreted as a proof of the non-existence of the impact of genetic heterogeneity in the population (*ibid*.). However, Gavrilov and Gavrilova (1991) presented also several proofs that such an interpretation would be too simplistic and showed that the genetic heterogeneity contributes to the observed variability of life spans mainly during the early stages of life. Therefore at higher ages the genetic heterogeneity has only a marginal role.

Vaupel (1997) also mentions the behavioral factors which could possibly influence the mortality pattern. Also other authors emphasize these behavioral factors, for example the altruism of the elderly studied not only in the social species but also in the nonsocial ones. Except for the altruism, in many social groups the older individuals are naturally leaders of that group because of longer time of their knowledge accumulation (Carey, Gruenfelder, 1997).

The effort to learn more about the mortality pattern is recently connected also with the research in genetics and the study of the DNA (Johnson, Shook, 1997). This research is focused on finding of some longevity-determining genes which are often called “gerontogenes” (*ibid*.).

Gavrilov and Gavrilova (1991) introduced three possible hypotheses which according to them can significantly influence the future development of mortality. The first of them is the “ecological crisis hypothesis” which is connected with the assumption of a fixed biological limit of the life span around age ca 100–110 years. This theoretical limit is supposed to be tied with the modern stressful and unhealthy lifestyle, pollution and destruction of ecology what are the factors that could cause the future stability of mortality rates despite the expected advances in medicine. The second concept is the “endogenous causes hypothesis” based on the work of Bourgeois-Pichat and his “concept of a temporary limit of mortality decline”. It supposes that medicine can almost eliminate the exogenous causes of death but it would take a long time. As a result this hypothesis expects that there is a biological limit to a reduction of mortality for each age. The third hypothesis they introduced is the “limited reliability of the organism hypothesis” emphasizing that an organism is a system with finite reliability. Formally this hypothesis is connected with the distinction of the total mortality into two components (background and age-dependent component – see Chapter 8 of this Thesis or Gavrilov, Gavrilova, 1979; 1991). According to this hypothesis there is always some risk of death at each age (Gavrilov, Gavrilova, 1991).

Many theories of ageing are briefly introduced by Harris (2009). When dealing the possible limits on human life expectancy, he mentions for example the theory of:

- Vitalism, what is a theory that could be traced back to the early Egyptian times. This theory supposes that the human body is endowed by some vital energy which is decreased by some illnesses, accidents, but also by ageing and by the time itself. As the human body loses its vital energy it ages and finally dies.

- Cellular mutation, what is a theory supposing the mutation of genes. As the human body grows old the more mutated genes it contains so as it is more vulnerable to ageing and death.
- The human machine theory seems to be similar to the previous one supposing that the failures of the body accumulate during its life and finally lead to death. It represents in fact the theory of limited reliability introduced by Gavrilov and Gavrilova (1991). Harris (2009) points out the similarity to the functionality of some machines, for example cars. Similar to the assumption of the accumulations of failures are also other theories called as the “burnout theory”, “the accumulation of cellular waste theory” or the “free radical theory”, all of them are traditionally studied by biologists rather than by demographers.
- Reproduction related theories derived often by application of reliable theories formulated for other species to humans. These theories are dealing the fact that human on average live much longer than is their reproduction period. Opinions about the reasons of this fact are often related to the behavioral factors mentioned already above (Vaupel, 1997; Carey, Gruenfelder, 1997). These theories are often discussed by biologists who mention the “redundancy” of the human body, what means that for that reason human may have some organs in pairs (for example two lungs etc.), so that when one of this two organs fails the other is sufficient for life (Harris, 2009).
- Programmed obsolescence theory is based on the assumption that there is a limit of life span (or period of functionality) given by the limit of reproduction possibilities of individual cells in the human body (Hayflick, 1965). This concept is often called the “Hayflick limit” after its discoverer Leonard Hayflick (Harris, 2009).
- The theory of hormone based ageing is more connected with the biological research and study of the impact of specific types of hormones which theoretically could influence the life span.

Based on the theories and information mentioned above (which is only a small sample of all the theories and assumptions related to the ageing process and mortality) it could be concluded that opinions are really divided into two or three groups (as was mentioned in the introductory part of this chapter). The first group suppose that there is a limit we are approaching to because of the (mostly biological) reasons and limitations represented by the theories like the theory of limited reliability, cellular mutations etc. Within this assumptions two different trends could be distinguished – one supposing that the limit is fixed, the other that this limit could be postponed mostly by the biological, medical or genetic research and inventions. The opposite assumption could be formulated as the belief that there is no limit of the human life span in fact, what is determined by the ongoing research, genetic engineering and improvements of the life style, health care, etc.

That this issue is widely interesting and natural to almost everyone could be proved by the number of public discussions of this topic, the raising interest of healthy life-style and environment (mainly in the developed countries). Representatives of various theoretical streams present their ideas and results of their research and publicly discuss the topic. As an example could be taken the research of de Grey (e.g. de Grey, 2004) who (based on the project SENS –

Strategies for Engineered Negligible Senescence[5]) supposes that important increase of the human life expectancy is not only possible but also probable through the reparation of the accumulated damage in the human body. His theory of the life lengthening is based above all on the genetic research and engineering. An opposite point of view is traditionally represented for example by Jay Olshansky (e.g. Olshansky, 2004). He mentions many scientists or scholars dealing with and believing in possible immortality in the history. Olshansky's point of view could be then expressed by his own sentence, "What do the ancient purveyors of physical immortality all have in common? They are all dead." (Olshansky, 2004). In this mentioned public discussion also one slightly compromise opinion could be remembered. For example gerontologist Steven Austad supposes that the life expectancy will continue its growth also without a groundbreaking medical or genetic invention and can probably reached values around 150 already by the half of the next century (BBC News, 2004).

While the mortality at lower ages is already at very low levels the research of demographers concentrate more and more to adults and to the oldest age groups. In practice, demographers when dealing with the mortality pattern at higher ages and its possible development in the future, traditionally use one of many so-called mortality laws which formally express the mortality pattern according to age. These mortality laws (often understand as less or more complicated analytical parametric functions) are applied to empirical data and are used for smoothing and extrapolation of these data mainly at higher ages where empirical data cannot be considered as reliable. Some of these functions will be briefly described in the Chapter 5 (introducing the SAS macro dealing with these functions), where also a method for parameter estimation will be introduced. There are many of these functions formulated in demography (for some of them see Gavrilov, Gavrilova, 1991 or Burcin *et al.*, 2010). For a researcher, there remains a problem which of these models to select and use in practice. Mostly it depends on his or her opinion about the mortality pattern. Gavrilov and Gavrilova (1991) formulated several principles for selecting the most appropriate mortality law. They mentioned for example to prefer models with theoretical justification or derived from theoretical hypotheses, the most universal laws as possible and models offering the best approximation with the fewest parameters.

3.3 Notation used in mortality models

In this brief sub-chapter the basic notation used traditionally in the field of mortality and mortality models will be described. The following notation will be used throughout the whole Thesis. For the expression of mortality levels and development there could be used three main statistical functions – the probability density function, survival function and hazard function. Those are standard functions used for the description of a random variable. In the case of mortality we can define a random variable (labeled as "X") representing the life span of an individual in our studied population. Then according to for example Wilmoth (1997) we can define the cumulative density function as:

[5] http://sens.org/sens-research

$$F(x) = P(X \leq x) = \int_0^x f(a)da,$$

where *f(x)* is the probability density function which describes the distribution of life spans in the population (related with the table function d_x or *d(x)* in the continuous form) and *x* stands for age. The cumulative density function expresses the probability that the life span of an individual is shorter than *x* years. The complement of this function to the value of one is the survival function which could be written as

$$S(x) = P(X > x) = \int_x^\infty f(a)da = 1 - F(x).$$

The survival function is related with the life table survival function, labeled as l_x / l_0 or *l(x) / l(0)* if we consider the continuous form. Then the hazard function could be expressed as (*ibid.*) as

$$h(x) = \frac{f(x)}{S(x)}.$$

This function expresses the probability density which is conditional on survival to that age. This function is traditionally in demography labeled as $\mu(x)$ what is the force of mortality (or intensity of mortality). The force of mortality could be simply derived from the standard table functions. Gavrilov and Gavrilova (1991, p. 41) derived the formula for the hazard function from the probability of dying (q_x), where

$$q_x = \frac{l(x) - l(x+1)}{l(x)},$$

when we consider the age interval (Δx) as equal to one year. Then it is possible to derive

$$\mu(x) = \lim_{\Delta x \to 0} \frac{l(x) - l(x+\Delta x)}{\Delta x * l(x)} = \frac{-\mathrm{d}l(x)}{l(x)\mathrm{d}x}.$$

Koschin (2002, p. 16) proved that the same equation could be derived in a similar way also for the age-specific mortality rates (m_x):

$$m_x = \frac{l(x) + l(x+1)}{l(x+1/2)}$$

when we consider the age interval (Δx) as equal to one year. Generally for any age interval Δx it could be written:

$${}_{\Delta x}m_x = \frac{l(x) - l(x+\Delta x)}{l\left(x+\frac{\Delta x}{2}\right) * \Delta x},$$

then

$$\mu(x) = \lim_{\Delta x \to 0} {}_{\Delta x}m_x = \lim_{\Delta x \to 0} \frac{l(x) - l(x+\Delta x)}{l\left(x+\frac{\Delta x}{2}\right) * \Delta x} = \lim_{\Delta x \to 0} \frac{-1}{l\left(x+\frac{\Delta x}{2}\right)} * \frac{l(x+\Delta x) - l(x)}{\Delta x} = \frac{-\mathrm{d}l(x)}{l(x)\mathrm{d}x}.$$

From the above written it is clear that when the considered age interval (Δx) goes limitedly to zero then there is no difference between age-specific mortality rate and the probability of dying. Both these functions approach the force of mortality.

3.4 Three the most important theories related to the limits of expected future improvements in mortality

Developed populations showed a significant increase in life expectancy during the 20th century. An obvious reaction to this fact was the consideration of demographers and other experts about the likelihood that there exists some limit value of life expectancy and its increase. Many approaches and theories dealing with the question of the existence of this limit of life expectancy were formulated. However, Wilmoth (1997) introduced a possible way of thinking about the potential limit. According to him, the theoretical limit must be describable and could be characterized through the mechanism by which influences the human life. In addition to that the limit could be characterized also through the way of its own manifestation. Based on it, it could be not only found but also quantified using demographic or statistical analysis. Just the way of manifestation of a potential limit is the objective of analysis of many demographers. Wilmoth (1997) defines three basic hypotheses: “limited-life-span hypothesis”, “limit-distribution hypothesis” and “compression-rectangularization hypothesis”.

Limited-life-span hypothesis is one of the most commonly reported or used in practice. It basically assumes the existence of a certain age, usually labeled as ω, which could no one of the population survive. In case of validity of this hypothesis it is theoretically possible to live up to the age of ω, but to survive it even for a single day would have been impossible (*ibid.*). For this reason the hypothesis has been criticized. For support of this theory two demographic arguments are used. The first is the assumption that the maximal achieved age in the human population does not increase in time and the second one is the assumption of the Gompertz-Makeham law as a sufficient expression of the dependence of mortality on age. Both these assumptions cannot be considered as correct, because as evidenced by the empirical data the highest achieved age does increase (Kannisto *et al.*, 1994 or Wilmoth, 1997). Similarly, Gavrilov and Gavrilova (1991) showed that even the validity of the Gompertz-Makeham law would not guarantee the existence of the limit of the attainable age. In addition, the weaknesses of the Gompertz-Makeham law for expressing mortality at the highest age have been demonstrated in several studies (Wilmoth, 1997; Koschin, 1999). Even on the basis of this evidence of invalidity of the assumptions, this hypothesis cannot be unambiguously rejected. It is theoretically possible that there exist some limit of longevity but that it lies at a much higher age than for which demographic data are obtainable for analysis (Wilmoth, 1997).

The limited-life-span hypothesis can be formally expressed as the existence of any age where the probability density function and the survival function is equal to zero. On the other hand the hazard function or force of mortality at this age goes to infinity. That means:

$$f(\omega) = 0$$
$$S(\omega) = 0$$
$$h(\omega) = \infty \, .$$

The limit-distribution hypothesis assumes the existence of some limit distribution of mortality by age, where the real data can not fall under its values. In this theory there are not, however, considered genetic differences or perhaps a theoretical change of this limit distribution caused for example by the environmental conditions or Darwinian selection (Wilmoth, 1997).

Attempts to express such a limit distribution are usually associated with estimates of intensity of mortality while some (usually exogenous) causes of death are excluded. Such analyses were carried out mostly by Bourgeois-Pichat, Kannisto or Olshansky. But these attempts neglect the fact that exogenous and endogenous causes are not always independent. It is also true that the influence of the exogenous causes changes with age (*ibid.*). Fries (1980) tried to find such a limit distribution by the extrapolation of the life expectancy. He concluded about the estimate of around 85 years. It corresponds to his results made under the assumption of the mortality compression (see below).

The compression-rectangularization hypothesis is often presented as a separate demographic issue. In general, the principle of the hypothesis could be expressed as a reduction of variance (variability) of ages at death in a given population over time (Wilmoth, 1997; Fries, 1980):

$$Var(X, t_1) > Var(X, t_2) > \cdots > Var(X, t_n),$$

where

$$t_1 < t_2 < \cdots < t_n.$$

The process of rectangularization or compression of mortality can be defined also by many other indicators. In relation to the existence of a possible limit for the length of life it is mentioned an assumption of this hypothesis that the compression reflects the convergence toward the potential maximal characteristic length of human life (Fries, 1980). However, as noted Wilmoth (1997), the only reduction in variability of ages at death is not the sufficient evidence for the existence of the limit of mortality or life expectancy. Along with the compression it can occur also a shift of the whole age distribution of deaths to higher ages (Wilmoth, 1997; Canudas-Romo, 2008). As a result, the compression or rectangularization of the survival curve would not express any evidence of the existence of some limit of human life. Nevertheless, the analysis of the process of rectangularization itself can play an important role in the analysis of mortality at the highest ages and therefore there is also an independent chapter devoted to this issue within this Doctoral Thesis (see Chapter 7).

3.5 Summary

This chapter could be taken almost as an introduction to the more analytical part of the Thesis. It presented some the most important theoretical approaches to mortality and its development in time. Questions related to ageing and human life span are natural to most of humans. Together with the improvements reached during the last decades also many new approaches revealed dealing with the possible future improvements of mortality.

Demographers could be divided according to their assumptions about the future mortality improvements into three main groups. While optimists (representing the “visionary” view) believe in significant rapid decreases in mortality (usually as a consequence of genetic engineering or important medical breakthroughs), pessimists (or “traditionalists”) assume that the limit of life span cannot be significantly postponed in the future. Some rather compromising view (shared by “empiricists”) is based above all on the past and contemporary development of mortality. As a result the empiricists expect some future improvements but not so significant in comparison to the optimists mentioned above.

Many scientists, not only demographers, are trying to find any clear signs of the future mortality development. How many scientists are studying mortality, almost so many approaches to this issue could be traced. Some of them work mostly in the theoretical field. Dealing with the behavioral factors, the theories attempt to explain the mortality pattern at higher ages. Other group of theories is based on the genetic research, study of cells or DNA. From the demographic point of view three important hypotheses were formulated which are tied to the way of manifestation of any hypothetical limit of human life span Wilmoth (1997) – the limited-life-span hypothesis, the limit-distribution hypothesis and the compression-rectangularization hypothesis. The compression of mortality (or rectangularization of the survival curve) became an almost independent analytical approach to mortality. It will be also the theme of one chapter of this Thesis (Chapter 7).

All the ideological streams or formulated theories mentioned above represent the background upon which all the analytical efforts arose. Many of those theories are based on the development of life expectancy as one of the essential demographic indicators. The life expectancy is the output of the life tables. Because of this reason, and because the life tables are one of the most important tools in demography, in the following part of the Thesis the theme of life tables will be opened.

Everybody wants to live forever,
but nobody wants to grow old

Jonathan Swift

Chapter 4

Current construction of the official life tables in the Czech Republic – weaknesses and alternatives

Life tables are an essential tool traditionally used in the analysis of mortality and aging. At the same time, it is the model that students of demography learn to construct usually as one of the first ones. However, the construction of life tables is not always clear and easy. A demographer while calculating the life table have to consider many aspects that can more or less influence the resulting values and so could be reflected in the values of the table functions.

The first factor that significantly influences the construction of life tables is the availability of necessary data. In the Czech Republic, there is no problem with data accessibility for the construction of the life tables at the national level as all traditional necessary inputs are published annually by the Czech Statistical Office[6]. The same situation holds usually also for other demographically developed countries. Aside from the potential problem with the input data, demographers have to tackle a number of other issues. They have to decide to which age a table will be constructed, whether the calculated mortality rates or probabilities of dying will be smoothed and extrapolated to higher ages, if so, what method of smoothing should be chosen, how to end the table, etc.

In this chapter, there the actual method of life table's construction in the Czech Republic will be described briefly. Together with this description also some important alternatives will be mentioned, that can be used by a demographer while calculating the table. Perhaps one of the most important factors is the choice of a suitable model of smoothing and extrapolation of the mortality rates or the probability of dying. Selected, probably the most frequent and most important, methods of smoothing and extrapolation are contained in an article that originated in the context of this Dissertation Thesis (Burcin *et al.*, 2010).

[6] http://www.czso.cz/eng/edicniplan.nsf/aktual/ep-4?opendocument

4.1 Current practice of the official life tables construction in the Czech Republic

A summary description of the construction of life tables can be found in its methodological Appendix (Czech Statistical Office, 2009b). The Czech Statistical Office publishes a comprehensive series of life tables from 1920 on its website[7]. All these tables are calculated in a uniform manner.

The calculation of the official life tables for the Czech Republic is done by the indirect method, so the probability of dying by age (q_x) is calculated from the age-specific mortality rates (m_x). According to the methodology of the Czech Statistical Office (2009b) thus could be written that

$$q_x = 1 - e^{-m_x}.$$

The probability of survival between age x and $x + 1$ (p_x) is calculated as a complement to the value of one from the probability of dying. From the formula of the probability of dying it could be derived easily that

$$p_x = 1 - q_x = 1 - (1 - e^{-m_x}) = e^{-m_x}.$$

The calculation of the probability of dying for the first year of life is different; it is expressed as the ratio of deaths at age 0 to the number of live births. Values of probability of survival and dying can be considered as the initial input for the construction of life table. If one evaluate their values according to age, it is possible to see significant deviations and random fluctuations. The standard approach is then to use some possible method of smoothing of these values before continuing the calculation of other table functions (Koschin, 2002). In the case of the Czech life tables, there the weighted moving average is used, which is applied to the ages from 4 years (Hartmannová, Fesenko, 1973; Czech Statistical Office, 2009b). Its construction is not complicated, and still the smoothing made by this calculation seems to be sufficient. The calculation can be expressed as follows (Hartmannová, Fesenko, 1973; Czech Statistical Office, 2009b):

$$q_x^{smoothed} =$$

$$= \frac{1}{315} * [105 * q_x + 90 * (q_{x-1} + q_{x+1}) + 45 * (q_{x-2} + q_{x+2}) - 30 * (q_{x-3} + q_{x+3})],$$

where $q_x^{smoothed}$ is the probability of dying between exact ages x and $x + 1$ smoothed by the weighted moving average. Due to the fluctuations of values at the highest ages it is important to find a suitable method of estimation of smoothed values also for these ages. In that case the extrapolation based on available and reliable data for younger ages is used (Koschin, 2002). There are many methods (or rather models) describing mortality at the highest ages formulated; they are partially described in more detail in the article of Burcin *et al.* (2010). In the Czech

[7] http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky_za_cr_od_roku_1920/$File/cr_ut_1920_2010.zip

demographic practice, it is traditionally chosen the usage of the Gompertz-Makeham formula (Czech Statistical Office, 2009b).

Using the Gompertz-Makeham formula makes it necessary to estimate its three parameters. This estimate has to be based on reliable data and the extrapolated values have to fit well the data for the highest ages. Given the need to estimate three parameters, it is necessary to use at least three empirical values. It is possible to base the parameters estimation on realistic (empirical) values for three selected ages (Pavlík *et al.*, 1986). In that case, the calculation is very easy, but it is necessary to mention the risk that one or more of these three selected ages used for the parameters estimation could be chosen inappropriately, it can be for example affected by any significant random variation. This can be avoided by using as many empirical values as possible and the parameters could then be estimated using for example the least squares method (Burcin *et al.*, 2010). Such a procedure may be technically challenging, especially if we do not have at least basic statistical software. A slightly compromise method of these two approaches could be the so-called King-Hardy method of parameters estimation (Fiala, 2005; Pecka, 1989). This approach is used also for the construction of the official Czech life tables (Hartmannová, Fesenko, 1973; Czech Statistical Office, 2009b). The King-Hardy method could be described in a simplified way as the calculation of three unknown parameters from the three given equations. The inputs are not the empirical values for three particular selected ages, but for three age intervals, which cover a sufficiently long age span. This calculation can be made relatively easily (Fiala, 2005) while almost eliminating the risk of inappropriate choice of the entrance ages. Detailed description of particular steps of the calculation used by the Czech Statistical Office is given in the methodological notes of the published life tables (Czech Statistical Office, 2009b). This method could be briefly described also according to Fiala (2005, pp. 36–39):

For the calculation it is necessary to choose 3 equally long following age-intervals, the first involved age will be labeled as x_0 and the length of these intervals is *k*. We consider the relation

$$\sum_{x=x_0}^{x_0+k-1} m_x = \sum_{x=x_0}^{x_0+k-1} \left(a + b * c^{x+1/2}\right) = G_1,$$

where traditionally m_x stands for the age-specific mortality rate. The sums on the left side of the equation above can be for each age-interval named as G_1, G_2 and G_3. Then the system of equations could be derived as

$$G_3 - G_2 = b * c^{x_0+k+1/2} * \left(c^k - 1\right) * \left(1 + c + \cdots + c^{k-1}\right)$$

$$G_2 - G_1 = b * c^{x_0+1/2} * \left(c^k - 1\right) * \left(1 + c + \cdots + c^{k-1}\right)$$

and after dividing of these equations we get

$$\frac{G_3 - G_2}{G_2 - G_1} = c^k,$$

where the parameter c can be easily calculated. Parameters b and a can be calculated from relations

$$b = \frac{G_2 - G_1}{c^{x_0+1/2} * (c^k - 1) * (1 + c + \cdots + c^{k-1})}$$

and

$$a = \frac{G_1 - b * c^{x_0+1/2} * (1 + c + \cdots c^{k-1})}{k},$$

where we can use the formula $1 + c + \cdots c^{k-1} = \frac{c^k - 1}{c - 1}$.

Neither the King-Hardy method is applicable without any problems. The problem of the need of subjective decisions when using this procedure has already pointed out Pecka (1989). Before the start of the calculation is necessary to make two essentially subjective decisions – first, to choose the initial age entering into the calculation (x_0) and, second, to choose the length of the three consecutive equally long intervals (k), from which estimated values are calculated. The Czech Statistical Office uses the age of 60 years as the first age entering the calculation and the length of the interval is considered to be equal to 8 years. The three intervals (outlined as described) thus cover ages 60–83 years. All the real (empirical) data for all these ages enter to the calculation process of the parameter estimation of the Gompertz-Makeham function (Czech Statistical Office, 2009b).

After the estimation of parameters of the Gompertz-Makeham function, this method is used also for the calculation of the values of probabilities of survival and dying for the highest ages. The Czech official life tables for year 2008 were finished at the age of 103 years (Czech Statistical Office, 2009a). Even the decision about this highest age in the table can be questionable and this highest attainable age could be chosen differently for various needs and practical use of the tables. Equally important is the way of ending the table. Hustead (2005) defined four basic ways of possible ending of the life tables:

1. "The Forced Method", when the life table is ended in some fixed age, where the probability of dying at this age is stated as to be equal to one. This procedure is used in the construction of tables by the Czech Statistical Office (Czech Statistical Office, 2009a). In that case, the probability of dying at the age of 102 years for men in the Czech Republic was for example in the year 2008 estimated as to be equal to 0.6376 and at age 103 it was equal to one (*ibid*.). Of course, this is a rather simplified ending of the table, which has not serious implications for the accuracy of such tables as the numbers of survivors at those maximal ages are already very small and often even zero.
2. "The Blended Method" is very similar to the first method mentioned above, also in this case the probability of dying at the highest age is considered as equal to one, but the values of the probability of dying at several previous ages are adjusted to gradually grew to the value of one (Hustead, 2005).

3. "The Pattern Method" is the least one of all those methods by its usage in practice (*ibid*.). In this method, there it is used exclusively the chosen method of extrapolation of the probability of dying or survival even at the highest ages (in the Czech practice, this would be the Gompertz-Makeham function). The highest age attainable in the life table is then chosen as such age, where the values of the method of extrapolation achieve one or are significantly close to one.
4. "The Less-Than-One Method" is similar to the previous one in that aspect that for no age there it is used the value of the probability of dying set to one. For the highest ages there are the values of probability of survival and dying calculated using the chosen extrapolation method and the table is ended at some selected highest age regardless of that the probability of dying at this age is less than one.

In the description of the construction of life tables in the Czech Statistical Office, there the last step of calculation left to be mentioned, and this is a natural transition from one type of smoothing (weighted moving average) to another (Gompertz-Makeham function). One can choose an age when this transition takes place and for several lower and higher ages it would be arranged a successive (weighted) change of these methods of smoothing. This age does not need to be the same for each year. In the Czech Statistical Office, there the method of the smallest difference is used. It means, the differences are calculated between the values obtained by both methods of smoothing and the age of transition is found as the age at which the difference is minimal (Czech Statistical Office, 2009b).

4.2 Disadvantages of the King-Hardy method for the parameter estimation of the Gompertz-Makeham function

As mentioned, the King-Hardy method for parameter estimation of the Gompertz-Makeham function is used in practice for the construction of the official life tables by the Czech Statistical Office (Czech Statistical Office, 2009b). This procedure eliminates the risk of inappropriate choice of the ages what could happen in the case of the calculation carried out only on the basis of three selected ages, as described for example in Pavlík *et al.* (1986). Nevertheless, even in the King-Hardy method a demographer have to make some subjective decisions that may significantly affect the final results of life tables.

The first step in the King-Hardy method of the parameters estimation of the Gompertz-Makeham function is the choice of two parameters – the first is the lowest entry age (x_0), which is the first age entering the process of parameter estimation, and the second is the length of the three consecutive intervals of equal length (k). The lowest age used for the Gompertz-Makeham function is traditionally chosen around the age of 60–65 years (Fiala, 2005). The Czech Statistical Office uses the lowest input age as equal to 60 years, and the length of three consecutive age intervals as 8 years (Czech Statistical Office, 2009b). Then the highest age entering into the calculation of the parameter estimates of the Gompertz-Makeham function is 83 years (because 60 + 3 * 8 – 1 = 83). It is this subjective choice of these two key values that can significantly affect the final results of life tables (Pecka, 1989), as will be shown below.

4.2.1 Presentation of possible consequences following from different choices of input values for the King-Hardy method of Gompertz-Makeham parameters estimation

To illustrate the impact of different subjective decisions about the input values (the lowest age and the length of age intervals) used for the estimation of the Gompertz-Makeham parameters a simple simulation prepared in the statistical software SAS 9.2 was used. For selected empirical data the parameters of the Gompertz-Makeham function were calculated for all the combinations of the lowest entering age (x_0) and the lengths of intervals (k) using the King-Hardy method. Based on the Gompertz-Makeham function and the parameters estimated, for each combination of x_0 and k the smoothed value of the force of mortality was calculated for the selected age and all these values were then compared. Ideally, if the King-Hardy method was not sensitive to subjective decisions of demographers and statisticians while its usage, the smoothed values for all combinations of input parameters were approximately the same for one selected age. The aim of this part of the text is to illustrate how the smoothed values can differ, if we change the values of the input parameters of the King-Hardy method.

The empirical data used for the model simulation were for females in the Czech Republic in 1997 but the results would be similar also for other years or populations. Parameter values of the King-Hardy method were considered for the lowest entering age as ranging from 50 to 71 years and for the length of the intervals from 1 to 10 years. For all the combinations of these parameters the estimation of the Gompertz-Makeham parameters was done. Based on these parameters the smoothed values for selected ages were calculated. Then these values were compared.

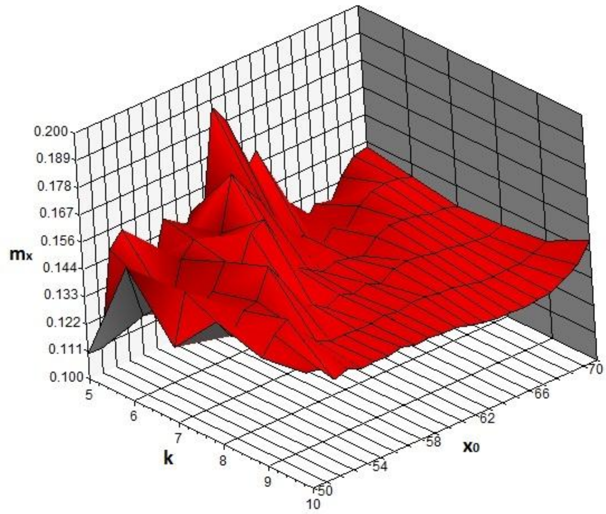
Results revealed that the procedure using the length of intervals shorter than ca 5 years leads to significantly fluctuating values, which is given by the fact that the Gompertz-Makemam parameters were estimated from only a small number of empirical values and they are also (because the considered age intervals directly follow at each other), concentrated into too narrow age interval. Therefore the results obtained for the length of intervals shorter than 5 years were omitted in the further analysis.

Based on the Figures 8–10 it is possible to directly compare the differences of smoothed values of age-specific mortality rates for different selections of the input parameters x_0 and k. Values vary more significantly for lower values of minimal entry age and shorter intervals. For longer intervals and higher ages the smoothed values fluctuate less, but still the differences are noticeable.

For the age of 85 years (Figure 8) the smoothed values still do not seem to be significantly variable, especially for the choice of the lowest entry age above 60 years and the length of the intervals at least 7 years. However, it should be noted that with increasing age the situation changes.

If the smoothed values are calculated for the age of 90 years (Figure 9), again there is noticeable a greater variation for the choice of lower input ages and shorter intervals. Nevertheless, even with "reasonable" choice of parameters of the King-Hardy method (i.e. the ages from ca 60 years and length of intervals of 7 years or more) it is possible to obtain relatively significantly varying smoothed values. For example, for the choice of the length of the intervals equal to 9 years, for our model data we can estimate a lower smoothed value if the lowest entry age is 68 or 70 years than if we take the first entry age as 62 years. Differences are evident from the Figure 9. If we compare the smoothed values for the age of 100 years, the differences are even more significant (Figure 10).

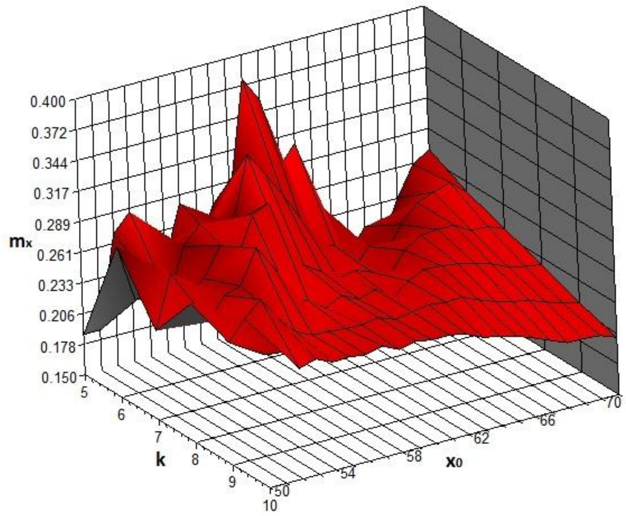
Figure 8: Smoothed values of the age-specific mortality rates (m_x) at age 85, Gompertz-Makeham function, King-Hardy method of parameter estimation where x_0 were considered from 50 to 71 years and the length of the intervals (k) from 5 to 10 years, Czech Republic, females, 1997



Note: Output from SAS 9.2 software

Source of data: Author's calculation based on Czech Statistical Office (2010–2012)

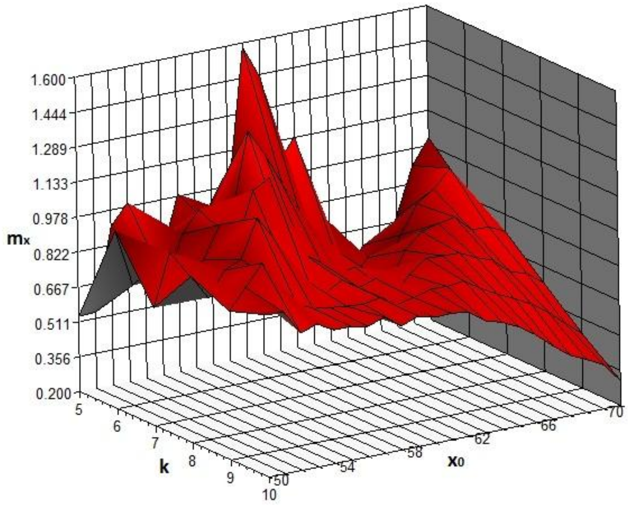
Figure 9: Smoothed values of the age-specific mortality rates (m_x) at age 90, Gompertz-Makeham function, King-Hardy method of parameter estimation where x_0 were considered from 50 to 71 years and the length of the intervals (k) from 5 to 10 years, Czech Republic, females, 1997



Note: Output from SAS 9.2 software

Source of data: Author's calculation based on Czech Statistical Office (2010–2012)

Figure 10: Smoothed values of the age-specific mortality rates (m_x) at age 100, Gompertz-Makeham function, King-Hardy method of parameter estimation where x_0 were considered from 50 to 71 years and the length of the intervals (k) from 5 to 10 years, Czech Republic, females, 1997



Note: Output from SAS 9.2 software

Source of data: Author's calculation based on Czech Statistical Office (2010–2012)

4.3 Summary

In this chapter the actual method of construction of the official life tables in the Czech Republic was introduced. It was mentioned that the Czech Statistical Office uses the Gompertz-Makeham method for smoothing and extrapolation of the empirical data at higher ages. Some the most important disadvantages of the currently used method of its parameters estimation were mentioned. Based on the brief examples shown above, it is clear that the current procedure of life tables' construction has some weaknesses, or at least there is an area where it is possible to consider a different approach to their construction. The currently used Gompertz-Makeham method and especially the method of estimation of its parameters showed to be the basic sources of possible changes of calculation.

Estimation of parameters of the Gompertz-Makeham function is currently done by the King-Hardy method in construction of the Czech official life tables. It is a simple variant of calculation, which bears virtually no demands on the equipment of statistical or other computing software. However, as shown, the price for this benefit appears to be relatively high, because the results obtained are very sensitive to the choice of input parameters – the first entering age and the length of the three consecutive age intervals. Therefore, the possibility of a potential modification of the currently used construction of life tables is addressed within this Thesis. The attention was mainly focused on various methods of smoothing of the probabilities of dying (or age-specific mortality rates). Some of these methods will be briefly introduced in the next chapter and some are described by Burcin *et al.* (2010).

When some parametric function is used for the smoothing and extrapolation of the force of mortality (it is not the aim of this Thesis to devote also to any semi-parametric or non-parametric methods) we have to find some suitable method of estimation of its parameters. It was mentioned that simple methods could be less accurate but their advantage is the clearness and the possibility to estimate the parameters without any specialized software. But the aim of this Thesis was also to present the possibilities of usage of statistical software in demography. Therefore, in the following chapter some of the described issues will be dealt and one of the most sophisticated statistical methods (the non-linear weighted least squares method) will be used for the parameter estimation. Also more methods of smoothing will be used, not only the Gompertz-Makeham function.

Death is not the end
Death can never be the end.

Death is the road.
Life is the traveller.
The Soul is the Guide
...
Our mind thinks of death.
Our heart thinks of life
Our soul thinks of Immortality.

Sri Chinmoy

Chapter 5

Smoothing of mortality rates using the SAS software

5.1 Introduction

SAS is modern statistical software and its possibilities of usages in demography are not yet fully discovered. There are many important advantages within the SAS software. First of all many inbuilt procedures could be used and so many difficult tasks could be solved quite easily. Secondly, the usage of SAS software enables to solve many problems and tasks in one moment or to repeat a procedure many times even for large data sets. As a result the solution is efficient and quick. Thanks to macros, the procedures or whole programs could be easily repeated or adjusted to another data. On the other hand, it must be pointed out that it is quite difficult to use the SAS software for someone who does not know even the basics of the SAS programming code (or language).

For the purpose of this work a simple macro (with several other sub-macros inbuilt) was prepared and it is part of the electronic Appendix of this Thesis. Also some model data are distributed with it. The main purpose of this macro is to produce the estimate of unknown parameters of several functions of mortality smoothing (mortality laws) and so to produce the first columns of the life table (that means the smoothed probabilities of dying). These values can be easily taken as the base for the life table construction according to selected method. Because this macro could be easily repeated for many calendar years or more populations, it is also possible to study the development of the parameter values itself. It can help to describe the mortality development as a whole. In the proposed macro a user can easily define own conditions (minimal and maximal ages used for the estimation, to select calendar years where the estimation process should be done, or select the function of mortality smoothing, etc.).

This chapter contains a short description of all the methods used in the proposed macro, a short description of the whole macro, and possibilities of its usage.

5.2 Basic description of the macro

Only parametric functions were implemented to the macro and so it could be possibly used in any of the following analyses. The first basic comparison of the most important functions, or rather mortality laws, was published in the Czech demographic journal *Demografie* (Burcin *et al.*, 2010). The method of estimation of the unknown parameters could be taken as an independent topic within the basic analysis of the mortality laws. In the proposed macro the weighted non-linear least squares method was used (will be described later). In some commonly used software, like Microsoft Excel or other types of spreadsheets, there the estimation of the unknown parameters using the mentioned method would be very complicated and almost impossible. Moreover, usage of such software for the estimation for more calendar years or for more populations is at least inefficient. That is why the statistical software (SAS) was chosen for the solution of this task. So that the usage of the macro was possible also for an inexperienced user of SAS, it will be described in detail within this chapter.

The main features of the macro program:

1) when the input data file respects the demands on its design, the macro could be submitted for data of many calendar years in one moment;
2) user can choose from several methods of mortality smoothing such as Gompertz, Gompertz-Makeham, Coale-Kisker and other functions (see below);
3) results are exported directly to Excel file where individual sheets contain particular years (populations). Name of the sheet contains not only the particular calendar year but also the abbreviation of the used method;
4) the estimation method minimizing the sum of weighted non-linear squares is implemented to the macro (more detailed description follows later in the text);
5) not only the estimated values of the parameters are exported but also the smoothed values of mortality rates with its confidence intervals, smoothed values of probabilities of dying and many other characteristics and results;
6) user of the macro has the possibility to choose the age-range for the estimation of the parameters and also the maximal theoretically attainable age – up to that age the smoothed values are calculated. Theoretically there are no limits for this age selection but it should be selected reasonably – the age-range should be width enough for the parameter estimation (at least 15 or better 20 years).

5.3 Methodological background

5.3.1 Life table construction at higher ages – usage of the laws of mortality

Because of the variability of the mortality rates at higher ages some methods of smoothing and extrapolation of the mortality rates are traditionally used. Available and reliable empirical data are used for the estimation of the unknown parameters in the selected function. Through this function the values of mortality rates (hazard function μ_x) are extrapolated also for those ages where the reliable empirical data are not accessible or do not exist at all.

There could be many of those models (or laws of mortality) presented and used in the calculation. Some of the most important ones are presented in the article of Burcin *et al.* (2010). The practice of the Czech Statistical Office is to use the Gompertz-Makeham function (Czech Statistical Office, 2009b) as was mentioned in the previous chapter. The same method is used also by some other national Statistical Offices, like in Slovakia or Estonia (European Commission, 2003), because the Gompertz-Makeham method is one of the best known one.

When the Gompertz-Makeham (or any other parametric method of smoothing and extrapolation) is used, then there arose the necessity of the estimation of the unknown parameters. Only reliable empirical data should be used in the calculation and in the same time the extrapolated values have to fit well the empirical values for the highest ages.

Several methods of the estimation could be used. One of the simplest ones usable for the estimation of the Gompertz-Makeham parameters is described in Pavlík, *et al.* (1986). It is the estimation of the three unknown parameters from the three empirical values of the age-specific mortality rates for three selected ages. The procedure is very easy but still there is the risk that one or more of these three selected values of mortality rates will reach extreme values in some way and the result could be biased due to that (as was already mentioned above, see Chapter 4).

Application of some more sophisticated methods could lead to more accurate results of the estimation procedure. One the basic methods frequently used in similar kind of needs is the method which minimizes the sum of differences between the empirical and estimated values (sum of the least squares; used for example in Burcin, *et al.* 2010). It could be more complicated to use such a method without some specialized statistical software.

Slightly compromise procedure is the King-Hardy method of estimation of the parameters. Briefly it could be described as the estimation of the three unknown parameters on the basis of the empirical mortality rates from three age-intervals (more details are in the Chapter 4 and in Fiala, 2005; Pecka, 1989). This method is used also for the construction of the Czech official life tables (Hartmannová, Fesenko, 1973; Czech Statistical Office, 2009b). Also this method has important disadvantages (see Chapter 4 or Pecka, 1989; Burcin, Hulíková Tesárková, 2011). When any compromises need not to be done and a demographer could use some suitable statistical software, not only the minimization of the least squares could be used but even more sophisticated and complicated methods.

In the proposed macro the non-linear regression is implemented where the weighted sum of the generalized (non-linear) least squares is used. The successful usage of the weighted sum of squares could be found also in the article of Wilmoth (1995).

5.3.2 Weighted generalized least squares method for the parameter estimation

It is reasonable to expect a changing variability of mortality rates with age (heteroskedasticity in data). In accordance to that, specifically constructed weights were used in the computational procedure in the introduced macro. The weights are taken as being equal to the reciprocal values of variance of the mortality rates with age.

The weights were constructed in accordance to the assumption that when the number of deaths is considered to be binomially distributed, then also the mortality rates are binomially distributed (or "relatively binomially distributed" as it was called by Gerylová and Holčík, 1988, p. 69). The variance of such distribution could be written as (*ibid.*):

$$\frac{\pi * (1-\pi)}{n},$$

where π symbolize the relative variable (in this case the mortality rate or the force of mortality) and n is the theoretical number of events (the theoretical number of deaths during the time interval is equivalent to the total number of people living during the same time interval). So we can rewrite the variance as:

$$\frac{m_x * (1-\mathrm{m_x})}{P_x},$$

where m_x is the mortality rate at age *x* and P_x is the population living at completed age *x* in the middle of the studied time interval. The reciprocal value of the variance at age *x* will be taken as the weight at that age in the estimation procedure. Because the method of estimation (generalized/nonlinear weighted least squares) is an iterative method, in each step (iteration) the weights are recalculated with the actual estimate of m_x (based on the actual estimate of the parameters). The same formula for the weights is introduced also by Koschin (1981) or Fiala (2005).

The Gauss-Newton method of estimation was used in the introduced macro. The Gauss-Newton iterative method "regress the residuals onto the partial derivatives of the model with respect to the parameters until the estimates converge" (SAS Institute Inc., 2009, p. 2926). The criterion of convergence has to be specified in the estimation procedure. In the macro attached to this work, the maximum change of parameter estimates as the convergence criterion was used. The iterations are said to have converged for `CONVERGEPARM = c` if

$$\max_j \left(\frac{\left|\beta_j^{(i-1)} - \beta_j^{(i)}\right|}{\left|\beta_j^{(i)}\right|} \right) < c,$$

where $\beta_j^{(i)}$ is the value of the *j*-th parameter after the *i*-th iteration.

When the parameters are estimated, values of the hazard function could be calculated easily only by substituting the parameter values into the equation of the selected model of extrapolation (below). For this purpose we have to describe the considered relation between the mortality rate (m_x) and the force of mortality (intensity of mortality μ_x). From the definition of those both terms it could be easily derived, that the intensity of mortality in the centre of the one-year age interval could be taken as equal to the mortality rate, and vice versa. Therefore, it is supposed that for all *x*, except the age zero, it holds (Thatcher, 1999):

$$\mu_{x+1/2} \approx m_x \,.$$

The probability of dying derived from the smoothed values of mortality rates were calculated as they usually are in the life table:

$$q_x = 1 - e^{-m_x} ,$$

Other life table functions could be then calculated easily from the column of probabilities of dying.

5.3.3 Description of the mortality laws used within this work

One of the most important advantages of the introduced macro is the possibility to use one or more different laws of mortality which are used as the parametric functions of mortality smoothing and extrapolation. Traditionally the exponentially or logistically increasing functions are used for the life table construction. Both these methods are represented in the group of models which are incorporated into the introduced macro.

The exponentially increasing models are represented by the Gompertz and Gompertz-Makeham formulas. The logistically increasing functions are represented by two formulas labeled as "Kannisto" and "Thatcher". The last two functions are slightly different – it is the Coale-Kisker model modified by Wilmoth (1995) to the quadratic function, and the modified Gompertz-Makeham function (proposed by Koschin *et al.*, 1998). The more detailed description of all the models could be found below:

Gompertz and Gompertz-Makeham functions

Too much description of those two (in demography probably the best known) mortality laws is not needed. The Gompertz function (Gompertz, 1825) is used in the form:

$$m(x) \cong \mu\left(x + \tfrac{1}{2}\right) = a * b^{(x+1/2)},$$

where a and b are the two unknown parameters which have to be estimated, x stands as usually for age. The Gompertz-Makeham function (Makeham, 1860) differs only by the usage of the constant a representing that part of the total mortality which does not change with age:

$$m(x) \cong \mu\left(x + \tfrac{1}{2}\right) = a + b * c^{(x+1/2)}.$$

As it was said already, the intensity of mortality (mortality rates) increases exponentially with age in both the mentioned functions. Because the estimated values of mortality rates approach the infinity with increasing age and because of the relationship

$$p_x = e^{-m_x} ,$$

the probability of survival approaches limitedly zero and the probability of dying approaches one for the highest ages in the Gompertz and also in the Gompertz-Makeham function if the intensity of mortality increase limitedly to infinity.

Modified Gompertz-Makeham function

The traditional Gompertz-Makeham function is based on the assumption that the rate of increase of mortality is constant with age and it models the age-related mortality by 3 parameters. Nevertheless, the empirical data show that this is not valid and at the highest ages the rate of increase of the mortality rate is very likely to slow down, so it is not constant, but rather a decreasing function. The model formulated on the basis of this mentioned assumption would contain one parameter more in comparison to the Gompertz-Makeham function, i.e. 4 parameters. The new parameter γ represents the mentioned fall in the rate of increase of mortality with age. This modified Gompertz-Makeham function could be in a form expressed as (Koschin *et al.*, 1998; Koschin, 1999):

$$\mu(x) = a + b * c^{x_0 + \frac{1}{\gamma} * ln[\gamma * (x - x_0) + 1]},$$

where $x > x_0$; a, b and c are the parameters of the traditional Gompertz-Makeham function, and $\mu(x)$ is the intensity of mortality at age x.

The age where the function starts to be applied, is denoted as x_0 and it is recommended to be chosen around the age of 85. Therefore, it is a model specifically focused only on the highest age groups. Ages entering into the calculation of the parameter estimates thus start at the age of around 85 years and end at ages where the empirical values of mortality rates begin to fluctuate significantly, i.e. slightly over 90 years (around ages 93 to 95 years). The smoothed values of the modified function resemble the course of the classic Gompertz-Makeham function (on which it is linked smoothly thanks to its construction), however, at the highest ages it does not grow so rapidly, what reflects the latest knowledge about the development of the oldest-old mortality (Koschin *et al.*, 1998; Koschin, 1999).

Kannisto and Thatcher functions

Just as the two previously mentioned basic functions (Gompertz and Gompertz-Makeham), differ the Kannisto and Thatcher functions only by the constant used in one of them.

The Thatcher function was proposed by Thatcher in 1999 (Thatcher, 1999), in the macro it is used in the form

$$m(x) \cong \mu\left(x + \frac{1}{2}\right) = \frac{a * e^{b * (x + \frac{1}{2})}}{1 + a * e^{b * (x + \frac{1}{2})}} + c\ ,$$

where a, b, c are the unknown parameters. The Kannisto formula does not contain the constant c again representing the component of mortality which is independent on age:

$$m(x) \cong \mu\left(x + \frac{1}{2}\right) = \frac{a * e^{b * (x + \frac{1}{2})}}{1 + a * e^{b * (x + \frac{1}{2})}}\ .$$

Both these functions are approaching the value of one with increasing age. As a consequence of that the probability of dying tends limitedly to $1 - e^{-1}$ = 0.632.

Coale-Kisker

This model was originally proposed as a relational one – the mortality rate at each age was measured in relation to the mortality rate at the age which was taken as the initial one in the model – usually 80 or 85. In the macro, there the model is used in its quadratic form derived by Wilmoth (1995):

$$m(x) = e^{a*x^2+b*x+c}\ .$$

5.4 Technical remarks – construction of the macro

The introduced macro is constructed from seven particular macros, each macro solves one of the 6 implemented functions plus one macro serves for running the whole process – this main "outer" macro reads the "setup row" where the demanded features of the calculation are defined by the user and automatically selects the proper "inner" macro solving the particular parameter estimation of the selected function.

Each "inner" macro starts by the import of relevant data – for the whole calculation only the exposure time (number of people living in the middle of the studied time period / year) and empirical values of death rates are needed. These values have to be available for at least one year for individual ages. The unified design of the input data is needed so that the correct data could be extracted from the data set within the macro. The input data set will be described later below.

The core of the macro is the estimation of the parameters. The procedure NLIN is used for this purpose. The general structure of the procedure used in the macro could be illustrated as show below:

```
PROC NLIN DATA=input_data_set MAXITER=&maxiter OUTEST=outest SAVE
CONVERGEPARM=1e-12;
BOUNDS a>&bounda;
PARMS a=&initial_value_of_a b=&initial_value_of_b;

IF
      (&minimal_age_of_estimation-1<x<&maximal_age_of_estimation+1)
THEN
      MODEL equation_of_the_model;

      IF (x<&maximal_age_of_estimation + 1) then
            _WEIGHT_ = description_of_the_weights_calculation;
      ELSE
            _WEIGHT_=0;

OUTPUT OUT=output_data_set PREDICTED=mxp PARMS=a b L95=lower
L95M=lowerM U95=upper U95M=upperM WEIGHT=w;

RUN;
```

The first row starts the NLIN procedure. Then the input data set is defined by the option DATA. The next option, MAXITER, defines the maximal number of iterations. In our macro the value of this parameter was selected as to be equal to 10,000. In most of the cases this value was not reached at all, usually only a few iterations are needed. In the macro the maximal number of iterations was stated as a macro variable (starts by the sign "&") so it is easy to change its value. The command OUTEST defines the data set for the estimates of parameters, in our macro the name of this data set is "Outest". The command SAVE is needed so that the final estimates of the parameters are saved to the OUTEST data set. The last option of the first row is CONVERGEPARM, its value is compared with the maximal change of the value of estimated parameters in each iteration. When the maximal change of the parameters is higher than this value, the next iteration starts, when the value of the maximal parameter change is lower than the defined value the estimation process is finished.

By the option BOUNDS some defined bounds of parameters could be set when needed. Again the value of the bound is defined as the macro variable so that it could be changed easily. Macro variables are also used for the definition of initial values of the parameters introduced by the command PARMS. The procedure showed out not to be very sensitive to these initial values. When these values are chosen within a reasonable range (where the real parameter value could be expected) then the results do not depend on the initial values.

The IF statement starts the main part of the procedure – it defines that the computation should be done only for the ages within the defined interval (user of the macro defines the minimal age of the calculation and the maximal age to which the estimated values should be produced). Then, after the statement MODEL, the description of the model is defined. For this particular purpose the model has to be stated like an equation. Then another IF process is started where the weights are defined. For values of ages higher than the defined maximal age (defined by the user) all weights are supposed to be equal to zero. Otherwise, the weights are calculated as described in the code, also the weights have to be defined in the equation form.

On the last row the parameters of the output are defined and labeled in the output data set. After the statement OUT= the output data set is defined. Then the predicted values of the mortality function are produced (labeled as "mxp") and the final estimation of the parameters (PARMS = a b). "L95" and "U95" specify variables that contain the lower/upper bound of an approximate 95 % confidence interval for an individual prediction. This includes the variance of the error as well as the variance of the parameter estimates. "L95M" and "U95M" specify variables that contain the lower/upper bound of an approximate 95 % confidence interval for the expected value (mean). In the described macro also the values of the weights from the last iteration are exported. The statement RUN; finishes the procedure and starts its processing when submitted.

Then the procedure SQL is used for the creation of the output table which is exported from the whole process to a defined Excel file, procedure SGPLOT produces the graphical output as defined.

Each macro for a particular function is finished by the export procedure:

```
PROC EXPORT DATA=&output_data_set_exported_to_excel
            OUTFILE= &adress_of_the_output_file
            DBMS=EXCEL replace;
     SHEET=&b&&function;
RUN;
```

On the first row of the procedure the data set with final results is specified, this data set was created in the SQL procedure and will be exported to Excel. This table is then exported to the defined address by the statement OUTFILE=. On the same row, there the Excel format is stated and the option REPLACE means that when a sheet with an already existing name is to be exported then the original (older) one would be replaced. For better orientation in the results the name of the sheet is defined in the form "Yxxxx", where "xxxx" signifies a calendar year (for example Y1980) – this is created by the calling of macro variable *b* ("&b"). After the 5 characters ("Y" and 4 numbers for the calendar year), there the abbreviation of the function would be added – each function is represented by one letter (the abbreviations are the same as at the beginning of the macro, each user has to select one abbreviation for the desired function.

When the currently (March 2012) latest version of the SAS software was released (SAS 9.3) it was necessary to prepare a slightly modified version of the macro. The latest SAS software in its 64-bit version uses a slightly different code of the procedure IMPORT and EXPORT. According to that a second macro with the same functions was prepared specifically for SAS 9.3, 64-bit users. This could be found also in the attachment. Except the way of importing and exporting data this second macro is the same as the first one.

5.5 How to use the macro – short manual for the user

The whole description of the usage of the macro below is prepared for the user who is not used to work with SAS software; a more experienced user would not need all the instructions, of course.

5.5.1 Input data file

The macro is designed for a standardized dataset, one such a model dataset is distributed (within this work) with some model data (in an Excel-format). The structure needed for the input data file is not complicated (see Figure 11). It should be an Excel file (xls-format) with two sheets:

1) sheet labeled as "Drates" containing age-specific death (mortality) rates, and
2) sheet labeled as "Exposures" with numbers of the exposed population (survivors to the middle of the year according to age or time exposure calculated in any possible way).

Both the sheets should have the same structure. The first column is expected to be labeled as "x" and contains the values of age ("0", "1", etc.). The last value has to be a number (not interval like "100+"). Other columns have to be labeled by the number of the particular calendar year and contain values of the death rates or numbers of survivors. It is no problem when the input data file contains more sheets – for example it is possible to use a sheet with numbers of deaths and from this sheet to create the sheet "Drates" by calculating the rates. When there is an unexpected structure of the input data sheet the macro will warn the used by an error-statement. It is possible to use the model data sheet. In this example, there is also the sheet called “Deaths” so that the values of death rates could be calculated. The sheet “Deaths” could remain as a part of the input data file, it will not be imported into SAS.

Figure 11: Example of the structure of the input data file

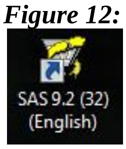
	A	B	C	D	E	F	G	H
1	x	1816	1817	1818	1819	1820	1821	1822
2	0	0,186986	0,181727	0,185714	0,196792	0,180915	0,18184	0,207284
3	1	0,046702	0,054247	0,061039	0,066216	0,05613	0,056687	0,060173
4	2	0,033928	0,038904	0,041661	0,045566	0,0393	0,041396	0,042095
5	3	0,022912	0,02705	0,028556	0,030022	0,026246	0,027983	0,02909
6	4	0,015995	0,018839	0,020295	0,021489	0,018371	0,019456	0,020213
7	5	0,013834	0,015245	0,016523	0,018145	0,015834	0,016255	0,016414
8	6	0,012102	0,012863	0,013993	0,015814	0,014054	0,014179	0,013975
9	7	0,01043	0,010745	0,011766	0,013601	0,012284	0,012121	0,011853
10	8	0,008907	0,008928	0,009821	0,011566	0,010563	0,010189	0,009885
11	9	0,007595	0,007492	0,008198	0,009657	0,008925	0,008444	0,008258
12	10	0,006286	0,006471	0,006964	0,007948	0,007366	0,006897	0,006979

Deaths / Exposures / Drates

5.5.2 Preparation of the macro

First of all, the folder “macro_store_smooth_32” or “macro_store_smooth_64” (for SAS 9.3 in 64-bit version users) distributed within this work should be copied to a home directory of the user. Also the program labeled as “SMOOTH32.sas” or “SMOOTH64.sas” should be saved to the home directory of the user and an input data file should be prepared. Then the SAS session have to be started. Here it will be described how to use the macro in the SAS software version 9.2 and we use the SAS BASE working space (not Enterprise Guide).

Figure 12: Icon of the SAS software



When the SAS session is started (icon is shown in the Figure 12) we need to open the code of the macro. When the Editor-window is activated, a user has to open the file “SMOOTH32.sas” or “SMOOTH64.sas” from his or her home directory (File – Open Program). The code will be opened in the Editor window and it is commented by many notes and instructions.

The first address which should be defined within the “Macro settings” is the directory where the folder “macro_store_smooth_32” (or “macro_store_smooth_64”) was stored before starting the SAS session.

In “Graphical settings” the folder have to be defined where the graphical outputs should be exported. Finally the input and output file should be defined.

The macro is started by some general instructions which should be read at least before the first usage of the macro (Figure 13). Then the user has to define addresses of several files as mentioned above (Figure 14).

Figure 13: General instructions of the macro

```
**********************************************************
* READ ME BEFORE THE FIRST PROCESSING                    *
*                                                        *
* Please read all the instructions below first.         *
*                                                        *
* Then fill the information which should be filled and  *
* submit the whole code or separatedly the particular   *
* parts of the code step by step.                        *
*                                                        *
* Please be sure that the input data file respects the  *
* needs for its design. If it does not, the programm    *
* wouldn`t be successful!                                *
*                                                        *
* Close the output file before submiting the code -     *
* without it the programm wouldn`t be successful!        *
*                                                        *
* It is highly recomended to read the log file after    *
* finishing the process.                                 *
**********************************************************
```

Figure 14: First part of the macro – definition of the input and output data file

```
*Macro settings:                                              ;
*Please fill the address where the file "macro_store_smooth_32" was saved and
submit the two rows (the code could be also submited as a whole);
libname smooth32 'g:\SAS\macro_store_smooth_32\';
options mstored mautosource spool SASMSTORE=smooth32;

*Graphical settings:                                        ;
*Please fill the address where graphical outputs should be stored and
submit the two rows (the code could be also submited as a whole);
ods html path="g:\SAS\pictures_store\";
ods graphics on;

*Please define the adress of the input data file;
%let adress_input="g:\SAS\data_example\FRA_F.xls";
*Please define the adress for the output file;
%let adress_output="g:\SAS\output_store\FRA_F_out.xls";
```

Figure 15: Second part of the macro – the setup row

```
*"SETUP ROW"
In the setup row (below) user has to input the values defining the output:
start = the initial (calendar) year for which the parameters should be estimated
stop = the last (calendar) year for which the parameters should be estimated
minimal = the initial age used for the estimation procedure
omega = the last age for which the smoothed values should be calculated
maximal = the last age used for the estimation procedure
function = the selected function of mortality smoothing
            - user can choose from this options (more detailed description
              of the functions could be found in the original Doctoral Thesis):
        K - Kannisto function
        M - Gompertz-Makeham function
        T - Thatcher function
        C - Coale-Kisker function
        F - modified Gompertz-Makeham function
        G - Gompertz function;
%smoothing (start=2000, stop=2005, minimal=30, omega=100, maximal=90, function=m);
```

The main part of the macro is the “setup row” (Figure 15). Several rows of instructions are written above the setup row, those should be read. The user has to fill the values in the setup row (in the brackets). The “start” value signifies the first calendar year for which the parameters of the model should be estimated. The “stop” value signifies the last calendar year for which the parameters should be estimated.

Three age-values should be defined then. The first one, it is the “minimal” value; it is the minimal age which should be used for the estimation. The “maximal” value signifies the maximal age which should be used for the estimation. By those two values the age interval entering the estimation process is defined. The empirical values of death rates at these ages are used in the estimation procedure. It is important that there are enough ages selected for this interval – at least ca 20 years should be used for the estimation of the parameters. Usually the more ages are in the selected interval, the better. On the other hand, only the ages where we suppose reliable data should be selected to this interval. Moreover, the ages where the selected function could be supposed to be valid should be used for the estimation of its parameters – that means that ages at least 25–30 years should be used because in lower ages the mortality development could be supposed to be different from the implemented mortality functions. For the modified Gompertz-Makeham function even higher ages should be selected, this model was originally developed for the highest ages (ca above 80 years, as mentioned above in the description of the model). Therefore, when the minimal age for this function would be selected as being equal to ca 80 years the length of the age interval used in the estimation process hardly could be 20 years.

The third value which has to be filled in is labeled as “omega”, what is the highest age for which the smoothed values are estimated. When the parameter values are estimated, then the smoothed values of the death rates are calculated using the estimated parameters and the formula of the selected model of smoothing. These smoothed values are calculated up to the age of “omega” and exported in the export data file.

The last value that should be defined is the selection of the function of smoothing and extrapolation. One can choose from six models (mortality laws) defined above – the Kannisto function, Gompertz-Makeham function, Thatcher function, Coale-Kisker function, the modified Gompertz-Makeham function and the traditional Gompertz function. There is an abbreviation defined for each of these functions (one letter – “K”, “M”, “T”, “C”, “F”, “G”). This letter should be written at the end of the “setup row” – see the Figure15 where the selected function is the Gompertz-Makeham function (`function = m`).

5.5.3 Setup the macro

When all the needed information is filled in the setup row the whole macro can be submitted. It is possible to submit the macro as a whole or to submit particular parts. The first way could be done easily when the Editor window is activated by clicking on the “submit” icon.

Figure 16: The “submit” icon in SAS

When one wants to submit particular parts of the macro one by one, it is necessary to select the particular part of the code (it is important to select the whole rows) and then again click on the “submit” icon (see Figure 16). This way of submitting is more practical for a user who wants to read the log output after each step of the code. The log output informs the user about the submitted process (part of the code). The log output is good to read after each submitting of the macro and it could be found in the “Log” window. After submitting the code, it could take several seconds (or even minutes) to process the code. The output file should not be opened when the process is running and it should be closed before the macro is submitted again.

5.5.4 Results

As a result the Excel file and graphical outputs in the png-format are created. In the output Excel file particular sheets contain results for particular selected calendar years and selected function (see Figure 17). The first 5 characters in the sheet name signify the calendar year (“Y2000” to “Y2005” in our example). The last character signifies the selected function (in our example it is “M” for the above selected Gompertz-Makeham function).

Figure 17: Sheet-names in the output data file from the SAS macro

37	65	0.007043	0.006927	0.004815	0.009039	0.006519	0.007335	0.006903
38	66	0.007639	0.007731	0.005505	0.009957	0.007287	0.008175	0.007701
39	67	0.008353	0.008644	0.006298	0.01099	0.008163	0.009126	0.008607
40	68	0.009383	0.009682	0.007217	0.012146	0.00916	0.010203	0.009635
41	69	0.010684	0.01086	0.008249	0.01347	0.010297	0.011423	0.010801
42	70	0.011604	0.012198	0.009402	0.014993	0.011591	0.012804	0.012123
43	71	0.012905	0.013717	0.010712	0.016722	0.013066	0.014369	0.013623
44	72	0.014511	0.015443	0.012229	0.018657	0.014745	0.016141	0.015324
45	73	0.015773	0.017403	0.013967	0.020839	0.016656	0.01815	0.017253

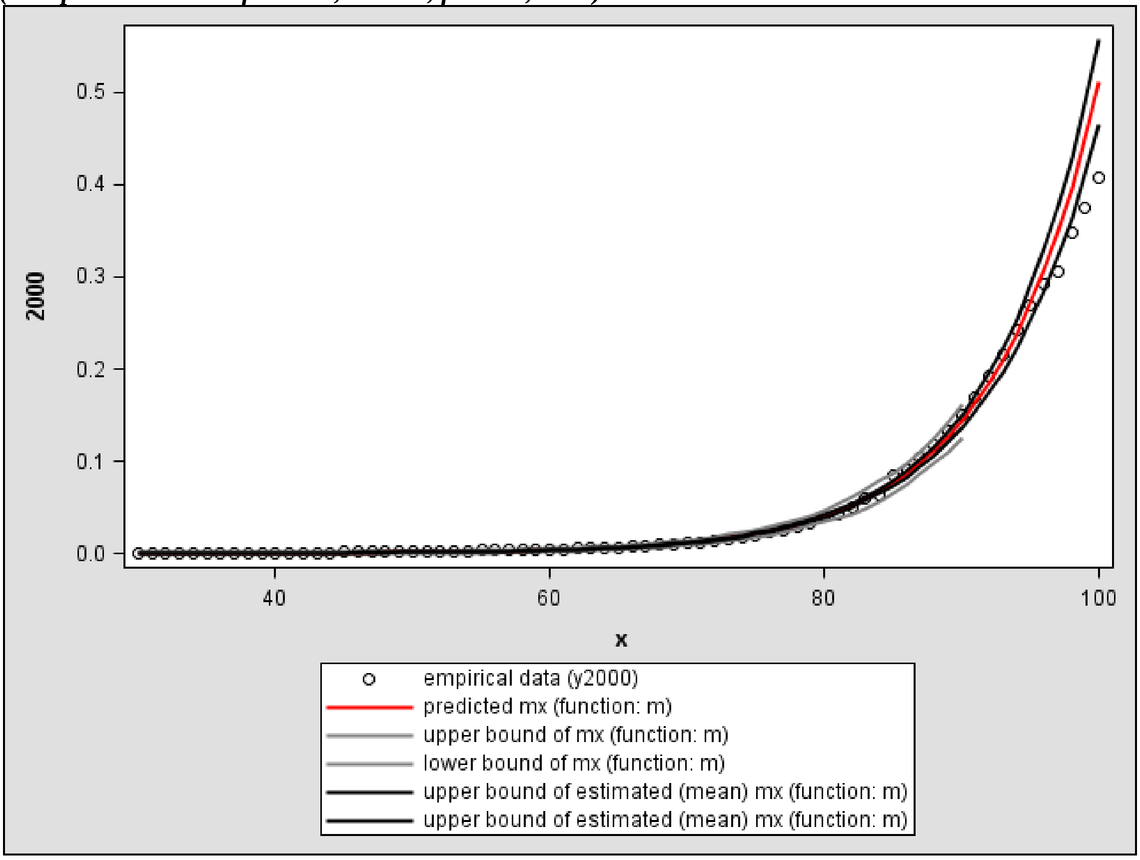
Y2000M | Y2001M | Y2002M | Y2003M | Y2004M | Y2005M

There is a unified structure of all the sheets in the output data file. In the first column, there is the age (labeled as “*x*” in accordance to the input data file). Then there the input data are repeated because there are empirical (imputed) death rates in the column named as “empirical_mx”. In the “predicted_mx” column, there are predicted values calculated through the usage of the formula of the selected function and the estimated values of parameters. Columns “lower”/“upper” and “lowerM”/“upperM” contain the lower/upper bounds of an approximate 95 % confidence intervals for the individual prediction (“lower”/“upper”) or for the expected value, the mean (“lowerM”/“upperM”).

The “predicted_qx” column contains the values of the probability of dying which could be used for the construction of the smoothed life table. These values are constructed on the basis of estimated values of the mortality rates. Using the lower and upper bounds of an approximate 95 % confidence intervals of the expected value (the mean) of mortality rates (“lowerM”/“upperM”) the bounds of the estimated values of probability of dying are calculated (“predicted_qx_lowerM”/“predicted_qx_upperM”). For comparison also the calculated empirical probabilities of dying are in the output file (“empirical_qx”). Its calculation is based on the empirical values of mortality rates imputed into the macro. Residuals (the difference between the empirical mortality rates and the estimated ones) are outputted in the column “residual_mx”. In the next column, “Px”, the original imputed numbers of survivors (or exposure time in general) are repeated. The final values of the weights are in the column named as “weight”. The next three columns contain the minimal age of the estimation, the highest age "omega", and the maximal age used in the estimation as were defined by the user in the "setup row" (see above). The last columns contain the final estimations of the parameters. The name of the column corresponds with the name of the parameter in the selected model as described above.

The graphical results are represented by one graph for each calendar year and the selected model. In these graphs there are the empirical values of mortality rates, the estimated values of mortality rates and upper and lower intervals of the predicted values and of the estimated values (the mean) of the mortality rates (see Figure 18). All the graphical outputs are stored in the defined folder and they are in the png-format.

Figure 18: Example of the graphical output from the macro without any further modifications (Gompertz-Makeham function, France, females, 2000)



Note: Output from SAS 9.2 software

Source of data: Author's calculation based on Human Mortality Database (2010)

5.6 Summary

In the previous chapters it was shown that not only the selected function of the method of smoothing and extrapolation can influence significantly the resulted life tables but also the method used for its parameter estimation. In the Chapter 4 it was illustrated for the Gompertz-Makeham function and King-Hardy method of parameter estimation. Within this chapter a more sophisticated method, the non-linear weighted least squares method, was not only introduced but also a tool for its calculation was prepared. In the attachment to this Thesis there a programming code for the SAS software could be found. It could be used for calculation of the parameter estimation for six selected mortality laws. The calculation could be easily repeated for more years or more populations in general. This macro is prepared in two modifications so as also users of the SAS 9.3 64-bit version could use it. This macro will be used also in several following parts of this Thesis.

Death is nothing at all.
I have only slipped away to the next room.
I am I and you are you.
Whatever we were to each other,
That, we still are.

Henry Scott Holland, Death Is Nothing At All

Chapter 6

Special issues related to life tables

In the previous chapters the theme of one of the most traditional demographic tools, the life tables, was opened. The most common way of life table construction was illustrated together with its weaknesses and also some selected mortality laws were introduced. It was show how the statistical software SAS could be used for the estimation of the parameters of mortality laws. In this chapter we will deal with some other special issues related to life tables.

In the first part, there the calculation of the age-specific mortality rates will be studied in detail so as the importance of the data quality could be shown and also the specifics of the period indicators. The topic will be illustrated on model and also on real individual data.

Methodology from the first part of this chapter will be used in the second one where the life tables will be used for the study of adult mortality according to education levels in the Czech Republic. In this part real individual data (Czech Statistical Office, 2011a) will be used.

6.1 Estimation of the exposure time for the age-specific mortality rates

Age-specific mortality rates are the basic input into the life table. That is the reason why it is worth dealing with that. Many demographers and users of demographic indices "are not aware of the fundamental difference between cohort and period dimension." (Luy, 2010b, p. 420). They are in fact often interested in cohort measures (according to the common interpretation of the used period indicators). Period measures are also difficult to interpret (Luy, 2010b).

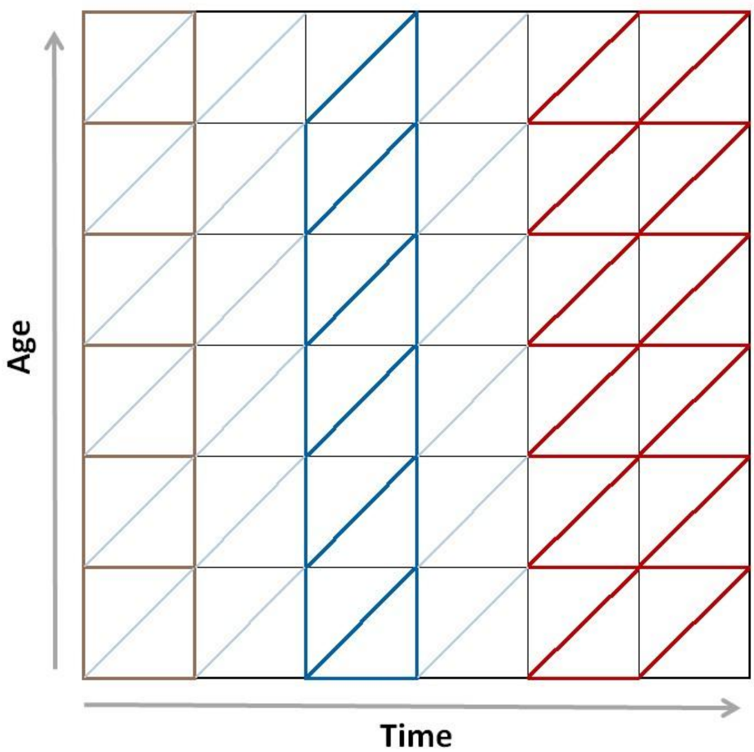
Usually there are three methods of the calculation of age-specific death rates defined in the demographic literature. The most frequently used one is the age-year-method based on the 3rd class of deaths (Wegner, 2010). Except of this traditional method also the cohort-year-method or age-cohort-method could be mentioned – the first of them is based on the 2nd class of deaths and the other is based on the 1st class of deaths (*ibid*.). Except of these three methods which result from the data classification we can distinguish other sub-methods how to calculate the exposure time or the denominator of the age-specific mortality rate in general.

As was said, traditionally the mortality rates are calculated in the period way so as numbers of deaths are summed for one age *x* and one calendar year *t* (representing the 3rd class of deaths). The standard and easy way of calculation of the period age-specific mortality rates is to divide the number of deaths at the completed age *x* and year *t* by the number of survivors in the middle of the year. The number of survivors in the middle of the year is often calculated as the average of the number at the beginning and at the end of the year (Wegner, 2010; Smith, 1992).

More exact way of the calculation divides the numbers of deaths by the amount of exposure to the risk (Preston *et al.*, 2001). But the calculation of the exposure could be quite difficult for real data. Calot and Franco (2002) mentioned and practically presented the importance of the selection of the proper and accurate method of the life table calculation. They recommended using the individual life durations where possible, mainly at higher ages.

In this chapter all the three basic methods of mortality rates will be used, for the 3rd class of deaths also the particular sub-methods will be show together with the differences which could be between the results. The calculation will be done for a model data first. First of all the situation will be illustrated for the 3rd class of deaths where it is probably the most complicated. Then the calculation will be repeated also for the 1st and 2nd classes of deaths. From the picture (Figure 19) it could be seen that all the classes of events could be formed by a proper arrangement of the elementary classes (triangles) of events.

Figure 19: Lexis diagram, 1st (red, right), 2nd (blue, in the middle) and 3rd (brown, left) classes of events (deaths) formed in a transversal view



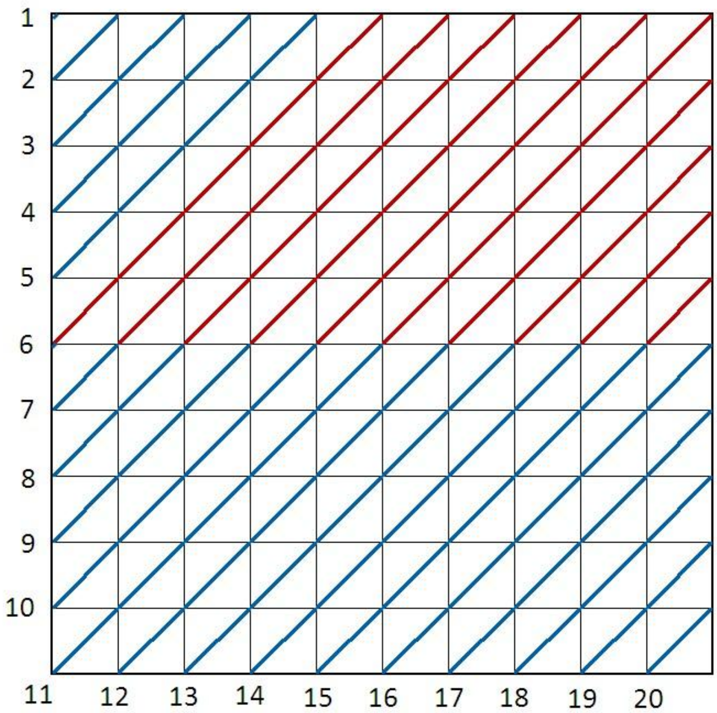
6.1.1 Calculation of the exposure time for model data

Let's define a model example (in the 3rd class of events) for the illustration of the calculation. Suppose, we have a model population where 100 persons live at the beginning of the year at completed age *x* and they could be divided into 10 particular sub-cohorts.

Then suppose that 10 persons die during the calendar year at completed age *x*. Five of them die during the first half of the year, five of them die during the second half of the year. However, it will be shown that in this case 97 persons could be alive in the middle of the year. How it is possible? And what the exposure time (the denominator of age-specific mortality rate) would be in this model example? For the example we will set two simplifying assumptions:

- all deaths occur at the same age of the person;
- all persons in the population were born at the beginning of the year (more exactly just after the beginning of the year) or *k * 1/10* of the year later, where *k* is constant equal to 1 to 9 (*k = 1, 2, …, 9*). This assumption corresponds with the fact that all the persons at the studied age could be divided into 10 sub-cohorts.

Figure 20: Lexis diagram for the model example, situation 1



Note: Numbers in the graph represent particular sub-cohorts living in the pictured rectangle (the 3rd class of events), i.e. during one year of age (vertical axis) and one year of time (horizontal axis)

The described situation is shown in the figure (Figure 20) where there are all the sub-cohorts (each born every 1/10 of the year) depicted. All the blue lines represent sub-cohorts containing 10 living persons, the red lines illustrate the sub-cohorts where one person deceased. The exposure time should be the sum of the time spent by all the people at the completed age *x* during time *t* (Preston *et al.*, 2001).

In the picture, there the period of one year is divided into 10 equally long time periods. Each 1/10 of the year exactly 10 people celebrates their x^{th} birthday. Exactly 0.5 year later one of those ten people dies. From that time only 9 people remains in the particular sub-cohort and grows older.

While we are interested in the exposure time, the denominator in the age-specific mortality rate, we have to sum up individual life durations during year *t* and age *x*. For easier description all the sub-cohorts appearing during year *t* at the completed age *x* got their numbers (see Figure 20). It could be seen also that according to our assumption there are not 10 sub-cohorts in the defined rectangular of the Lexis diagram, but there 20 cohorts appear during the year. Ten of them enter the rectangular by their x^{th} birthday and ten of them enter the rectangular at the beginning of the year when they were already *x* years old (i.e. 10 of them are members of the cohort $t – x$ and 10 of them are members of the cohort $t – x – 1$, these two groups of events form the two Lexis triangles in the diagram). Now we theoretically can sum up all the individual life durations and calculate the denominator of the age-specific mortality rate. The exposure time could be estimated for all the sub-cohorts as follows:

Sub-cohort 1: People in the sub-cohort 1 appear during year *t* at the age *x* only for a short time period – these people are born just after the beginning of the year, so just after the beginning of the calendar year *t* they celebrate their $x + 1^{st}$ birthday and so leave the rectangle. The time spent by those people in the pictured rectangle limitedly goes to **zero**.

Sub-cohort 2: Those people would celebrate their $x + 1^{st}$ birthday 1/10 of the year after the beginning of the year *t*, so they spend exactly **1/10** of the year in the rectangle. The time spent in the year and age could be read on the horizontal (time) axis.

Sub-cohort 3: Those people would celebrate their $x + 1^{st}$ birthday 2/10 of the year after the beginning of the year *t*, so they spend exactly **2/10** of the year in the rectangle.

Sub-cohort 4: Those people would celebrate their $x + 1^{st}$ birthday 3/10 of the year after the beginning of the year *t*, so they spend exactly **3/10** of the year in the rectangle.

Sub-cohort 5: Those people would celebrate their $x + 1^{st}$ birthday 4/10 of the year after the beginning of the year *t*, so they spend exactly **4/10** of the year in the rectangle.

Sub-cohort 6: Those people would celebrate their $x + 1^{st}$ birthday 5/10 of the year after the beginning of the year *t*, so they spend exactly 5/10 of the year in the rectangle. But one of this sub-cohort dies just after the beginning of the year (in time limitedly approaching zero), so only **9 persons would live 5/10** of the year in the studied time and age interval, **1 person would live** only for a short time interval (the time interval limitedly approaching **zero**).

Sub-cohort 7: Those people would celebrate their $x + 1^{st}$ birthday 6/10 of the year after the beginning of the year *t*, so they spend exactly 6/10 of the year in the rectangle. But one of this sub-cohort dies 1/10 of the year after the beginning of the year (life duration = 1/10), so only **9 persons would live 6/10** of the year in the studied time and age interval, **1 person would live 1/10 of the year**.

Sub-cohort 8: Those people would celebrate their $x + 1^{st}$ birthday 7/10 of the year after the beginning of the year *t*, so they spend exactly 7/10 of the year in the rectangle. But one of this sub-cohort dies 2/10 of the year after the beginning of the year (life duration = 2/10), so only **9 persons would live 7/10** of the year in the studied time and age interval, **1 person would live 2/10 of the year**.

Sub-cohort 9: Those people would celebrate their *x + 1st* birthday 8/10 of the year after the beginning of the year *t.* Only **9 persons would live 8/10** of the year in the studied time and age interval, **1 person would live 3/10 of the year**.

Sub-cohort 10: Those people would celebrate their *x + 1st* birthday 9/10 of the year after the beginning of the year *t.* Only **9 persons would live 9/10** of the year in the studied time and age interval, **1 person would live 4/10 of the year**.

Sub-cohort 11: Those people would celebrate their *x + 1st* birthday just after the beginning of the year *t + 1,* so they spend almost a whole year in the studied time and age interval (without a short time limitedly approaching zero). However, only **9 persons would live the whole year** in the studied time and age interval, **1 person would live 5/10 of the year**.

Sub-cohort 12: Those people would celebrate their *x + 1st* birthday during the year *t + 1,* but they entered the year *t* 1/10 of the year after the beginning of the year *t*, so they spend 9/10 of the year in the studied time and age interval. Only **9 persons would live 9/10 of the year** in the studied time and age interval, **1 person would live 5/10 of the year**.

Sub-cohort 13: Those people would celebrate their *x + 1st* birthday during the year *t + 1,* but they entered the year *t* 2/10 of the year after the beginning of the year *t*, so they spend 8/10 of the year in the studied time and age interval. Only **9 persons would live 8/10 of the year** in the studied time and age interval, **1 person would live 5/10 of the year**.

Sub-cohort 14: Only **9 persons would live 7/10 of the year** in the studied time and age interval, **1 person would live 5/10 of the year**.

Sub-cohort 15: Only **9 persons would live 6/10 of the year** in the studied time and age interval, **1 person would live 5/10 of the year**.

Table 1: Calculation of the exposure time, model example

A	B	C	D	E	F = B*C + D*E
Sub-cohort	Number of survivors	Life duration of 1 survivor	Deaths	Life duration of 1 deceased person	Total exposure time
1	10	0.0	0	–	0.0
2	10	0.1	0	–	1.0
3	10	0.2	0	–	2.0
4	10	0.3	0	–	3.0
5	10	0.4	0	–	4.0
6	9	0.5	1	0.0	4.5
7	9	0.6	1	0.1	5.5
8	9	0.7	1	0.2	6.5
9	9	0.8	1	0.3	7.5
10	9	0.9	1	0.4	8.5
11	9	1.0	1	0.5	9.5
12	9	0.9	1	0.5	8.6
13	9	0.8	1	0.5	7.7
14	9	0.7	1	0.5	6.8
15	9	0.6	1	0.5	5.9
16	10	0.5	0	–	5.0
17	10	0.4	0	–	4.0
18	10	0.3	0	–	3.0
19	10	0.2	0	–	2.0
20	10	0.1	0	–	1.0
Total	190	–	10	–	96.0

Sub-cohort 16: Nobody from this sub-cohort would die during year *t* at age *x*, so **all 10 people would spend 5/10 of the year** in the studied time and age interval.

Sub-cohort 17: Nobody would die during year *t* at age *x*, so **all 10 people would spend 4/10 of the year** in the studied time and age interval.

Sub-cohort 18: **All 10 people would spend 3/10 of the year** in the studied time and age interval.

Sub-cohort 19: **All 10 people would spend 2/10 of the year** in the studied time and age interval.

Sub-cohort 20: **All 10 people would spend 1/10 of the year** in the studied time and age interval.

So according to the previous detailed calculations we can estimate the total exposure time (number of exposed to the risk of dying) as could be seen in the Table 1.

Or we can develop a formal equation for population pictured in Figure 20. There are 40 red parts of the year (lived by the sub-cohorts with only 9 persons) and there are 60 parts of the year where all 10 persons of the sub-cohorts were alive:

*exposure time = (40 / 10) * ((no. of people at the beginning of the year / no. of sub-cohorts) – no. of people dying in each sub-cohort) – (60 / 10) * (no. of people at the beginning of the year / no. of sub-cohorts)*

For the population shown in the Figure 20 we can then calculate exposure time according to the formula above as:

*exposure time = (40 / 10) * ((100 / 10) – 1) – (60 / 10) * (100 / 10) = 4 * 9 + 6 * 10 = 96*

So the mortality rate for age *x* and calendar year *t* would be equal to 10 / 96 because there are 10 deaths in the defined rectangle and the exposure time was estimated as 96.

These results answer the questions set at the beginning of this part of the chapter. When we have 100 persons at the beginning of the year *t* at completed age *x* and 10 persons die during the year and age, so there are not 90 persons at the end of the year at age *x*, as would be the standard answer, but there could be 95 persons at the end of the year at the completed age *x* because equally numerated sub-cohorts enter the rectangle during that year. Of course, it is the effect of the selected 3rd class of events and of our simplifying conditions set above. The example should be taken only as an illustration of the main idea.

Let's continue briefly with a similar model population where the only difference is in the age at which people die – now they die at the age *x + 0.7.* Again all could be seen in the picture in the Figure 21.

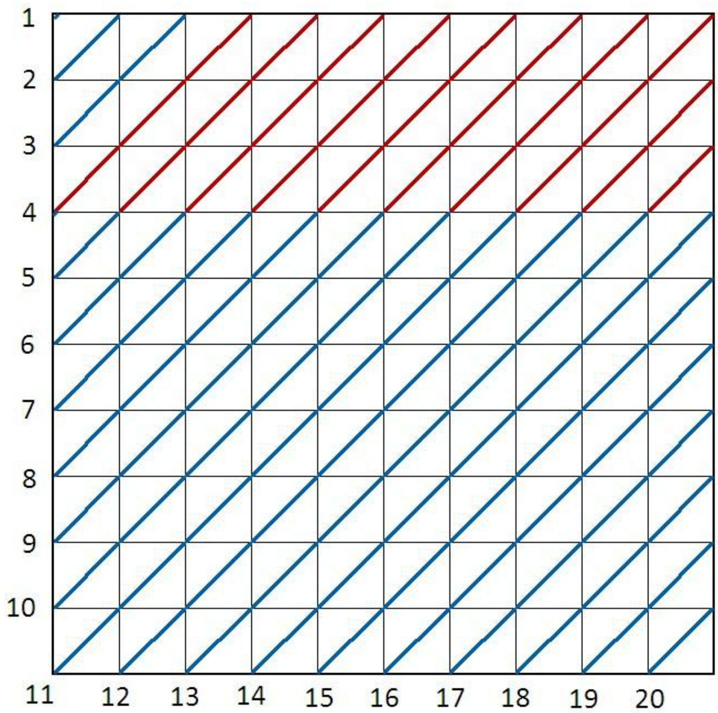
In this case the exposure time could be calculated as (using the same formula as in the previous example, now there are 27 parts of the year or rectangle where only 9 persons remains alive and 73 parts where 10 persons live):

exposure time = (27 / 10) * ((no. of people at the beginning of the year / no. of sub-cohorts) – no. of people dying in each sub-cohort) – (73 / 10) * (no. of people at the beginning of the year / no. of sub-cohorts) =

= (27 / 10) * 9 – (73 / 10) * 10 = 97.3

The age-specific mortality rate would be in this case equal to 10 / 97.3

Figure 21: Lexis diagram for the model example, situation 2



Note: Numbers in the graph represent particular sub-cohorts living in the pictured rectangle (the 3rd class of events), i.e. during one year of age (vertical axis) and one year of time (horizontal axis)

These results show another interesting fact: when we have 100 persons at the beginning of the year *t* at completed age *x* and 5 of them die during the first half of the year and 5 of them during the second half of the year, then at the end of the year there still could be 97 persons alive and in the middle of the year there could also be 97 persons alive in our model example.

Now it is clear how it is possible that we have 97 persons alive in the middle of the year, although 5 persons from 100 at the beginning of the year deceased during the first half of the year and another 5 persons died during the second half of the year, what was the original question set at the beginning of this part of the chapter.

Of course, these simple model calculations were not shown only because it might be interesting but the reason was to show the idea and the detailed method of calculation of the exposure time when individual data are available. The method will be used in the next part where differently calculated age-specific mortality rates will be compared and also this method will be practically used in further parts of this Thesis.

6.1.2 Real data for the calculation of the exposure time

In the previous part of the chapter, it was shown how it is possible to calculate the exposure time needed as a denominator for the calculation of the age-specific mortality rates in the case of the 3rd class of events. It was shown that the standard way of calculation when the number of deaths is simply divided by the number of survivors in the middle of the year (usually calculated as the average of the numbers of survivors at the beginning and at the end of the year) could lead to more or less different results in comparison with the calculation where all individual life durations are supposed. Of course the second method (individual life durations) is more complicated and data-demanding then the first one.

Very often it is almost impossible to calculate the individual life durations because we do not have the individual data for the studied population. But even if we have individual data about the persons deceased during the year almost never we have individual data about the survivors. These data could be taken only from the population census and population census is usually held only once in 10 years. That is why, for the real population we can develop some compromise method where the life duration of the survivors is estimated as the average of the survivors at the beginning and at the end of the year and the life durations of the persons deceased during the year can be calculated from the individual data. We will illustrate this compromise method on real data for the Czech Republic in 2010.

Because we would like to show what the differences of various calculations could be for the age-specific mortality rates we will consider above all the older ages where less people live and die because it could be supposed that for higher ages the difference should be more significant. We will compare the age-specific mortality rates from the standard method of calculation (number of deaths divided by the number of survivors in the middle of the year estimated by the Czech Statistical Office (2011b) or calculated as the average of the numbers of survivors at the beginning and at the end of the year) and from the slightly compromise method (where the number of deaths is divided by the estimated exposure time calculated from the individual data for people deceased during the year and estimated as the average of persons at the beginning and at the end of the year for survivors). For the analysis the ages from 80 to 99 years were used. We still suppose only the 3rd class of events.

First of all we will use the most detailed method – using the individual data and calculating the individual life durations. In the nominator of the age-specific mortality rate there is the total number of deaths that occur at the age x during year 2010. In the denominator there is the estimate of the exposure time. The exposure time was estimated as a sum of the exposure time of survivors and exposure time of persons deceased during the year at the completed age x.

The exposure time of survivors could be in the ideal case estimated from the individual lengths of life of the survivors but we have no individual data about those persons who do not die during the studied year. From the pictures in the first part of this chapter it could be seen that on average a person who do not die in the studied rectangle would spend 0.5 year alive at age x and year t (2010). That means that all the persons, who are alive at the end of the year spend on average 0.5 year in the rectangle. But within the number of people who are alive at the beginning of the year there are also some of those persons included who would die during the year. But not all of those deceased during the studied year are alive at the age x at the beginning of the year. As was already mentioned, there could be distinguished two groups of persons (cohorts) deceased in the year t and at the age x. For the estimation of the exposure time of survivors only those who entered the rectangle at the beginning of the year (they celebrated their

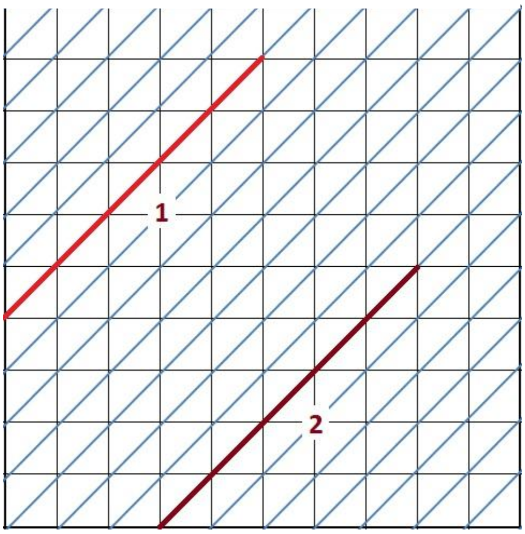
x^{th} birthday already during year *t – 1* and are members of the cohort born in the year *t – x – 1*, in the Lexis diagram these persons could be found in the upper triangle of the considered rectangle) have to be subtracted from the number of persons alive at the beginning of the year. So the estimate of the exposure time of survivors could be expressed as:

*Exposure time of survivors = (no. of survivors at the beginning of the year – no. of deaths in the upper triangle in 2010) * 0.5 + no. of survivors at the end of the year * 0.5 = (no. of survivors at the beginning of the year + no. of survivors at the end of the year – no. of deaths in the upper triangle in 2010) * 0.5*

The exposure time of deceased people at the completed age *x* and calendar year *t* could be estimated as a sum of exposure times for the two elemental classes (triangles):

a) of those who entered the rectangle at the beginning of the year (they celebrated their x^{th} birthday already during year *t – 1*), in the rectangle illustrated in the Figure 22 is one of those shown and marked by the red line and number 1 (upper triangle);
b) of those who entered the rectangle during the calendar year by their birthday. In the picture (Figure 22), there one of these is marked by the brown line and by the number 2 (lower triangle).

Figure 22: Representatives of two cohorts in the Lexis diagram of the 3rd class of events



For the first group of people the individual life durations (number of days they lived in the studied calendar year) are calculated simply as the difference of the date of death and the beginning of the year.

For the second group of people there the individual exposure time is the time between their x^{th} birthday and the date of death, that means that we had to calculate the number of days a person spend in the calendar year *t* at the completed age *x*. In the picture (Figure 22) both persons shown have exactly the same life duration, both spend 0.5 of the year in the rectangle.

When all the individual lengths of life are calculated in the described way then simply the sum of them is added to the estimated exposure time of survivors described above. In this way it is possible to obtain the total exposure time (the denominator of the age-specific mortality rate). We will mark the rates calculated in this way by the number “I”. That means that the age-specific mortality rate is calculated as:

$$m_{x,t}^{I} = \frac{D_{x,t}}{ET_{x,t}} = \frac{D_{x,t}}{ET_{x,t}^{survivors} + ET_{x,t}^{deceased}}$$

where $D_{x,t}$ is the number of deaths at the completed age *x* during the calendar year *t*, $ET_{x,t}$ is the exposure time lived by all the persons at the completed age *x* during the calendar year *t* (i.e. the sum of the exposure times of deceased and survivors for the lower and upper triangle in the considered class of events), $m_{x,t}^{I}$ is the age-specific mortality rate calculated in the first described way.

Other method of the age-specific mortality rate calculation is already more simple and more frequently used one. In the nominator there remains the total number of deaths. In the denominator there the exposure time is estimated by the number of persons who are alive in the middle of the year. Number of persons living at the completed age *x* in the middle of the calendar year *t* is traditionally published by the statistical offices in particular countries. Rates calculated in this way are marked in the graphs by the number “II”.

$$m_{x,t}^{II} = \frac{D_{x,t}}{P_{x,1.7.t}}$$

where $D_{x,t}$ is the number of deaths at the completed age *x* during the calendar year *t*, $P_{x,1.7.t}$ is the number of people alive in the middle of the year *t* at the age *x*.

And finally we will calculate the standard age-specific mortality rates, marked by the number “III”. In the nominator of the rate there is the total number of deaths that occur at age *x* and during the year 2010. In the denominator, there is the number of survivors estimated as the average of persons living at the beginning of the year and at the end of the year *t*.

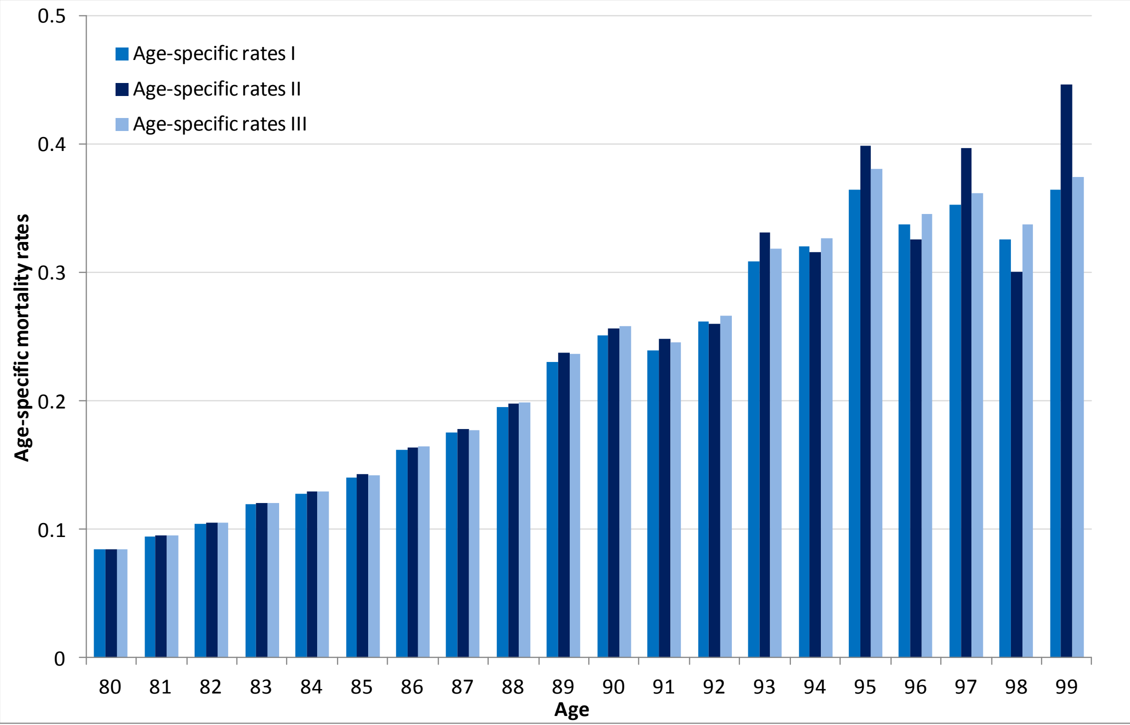
$$m_{x,t}^{III} = \frac{D_{x,t}}{(P_{x,1.1.t}+P_{x,31.12.t})/2}$$

where $D_{x,t}$ is the number of deaths at the completed age *x* during the calendar year *t*, $P_{x,1.1.t}$ is the number of people alive at the beginning of the year *t* at the age *x* and $P_{x,31.12.t}$ is the number of people alive at the end of the year *t* at the age *x*.

For the real data (Czech Republic, males and females, 2010) the age-specific mortality rates were calculated for the 3rd class of events in all three presented manners. The aim is to compare the results, show the differences and derive some conclusion about the importance of the data quality and detail.

From the pictures (Figure 23, 24) it could be seen that there is small difference in the values of the rates according the way of calculation for both sexes. But the differences become more significant with growing age. It is the consequence of lower numbers of survivors and deaths at higher ages and also of not a uniform distribution of deaths in the studied age- and time-interval. The second mentioned fact could be seen through the Lexis diagram again. We will consider only the highest ages where there are not many deaths and where the three calculated age-specific mortality rates are the most different.

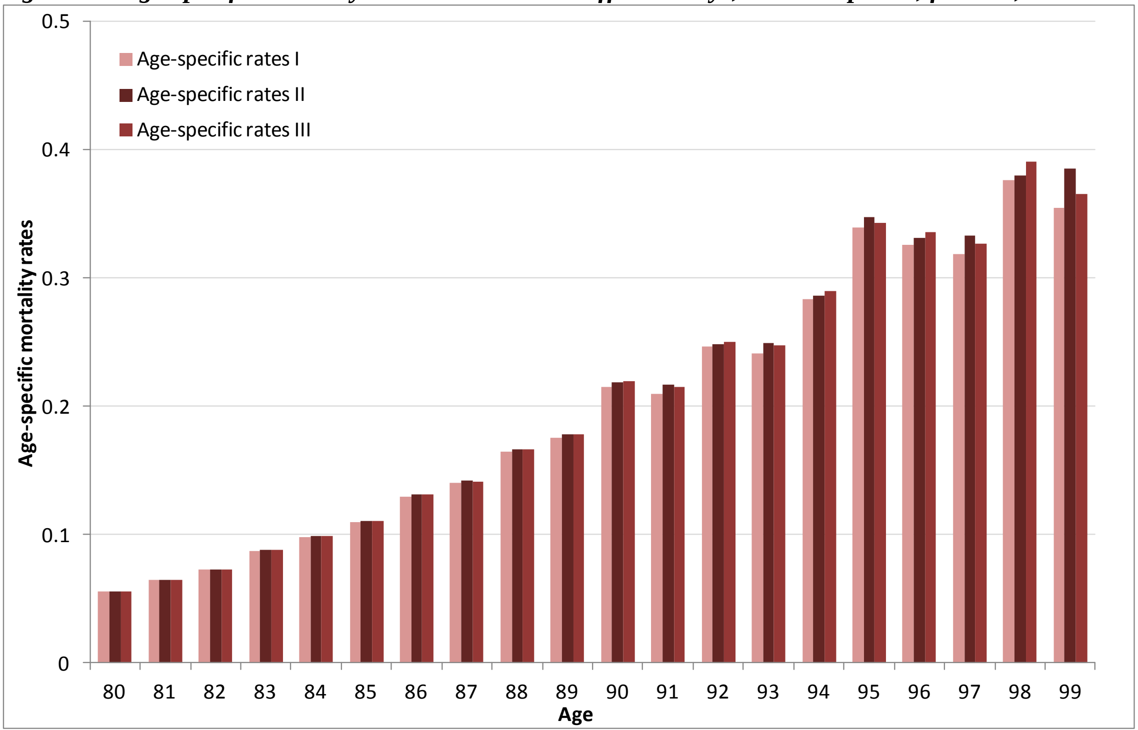
Figure 23: Age-specific mortality rates calculated in different ways, Czech Republic, males, 2010



Note: "Age-specific rates I" are rates calculated from the individual life durations at age *x* and in the calendar year *t*, "Age-specific rates II" are calculated using the numbers of living in the middle of the year as the denominator, "Age-specific rates III" are calculated with the average of the numbers of survivors at the beginning and at the end of the year in the denominator (for more details see the text of this chapter).

Source of data: Author's calculation based on Czech Statistical Office (2011a)

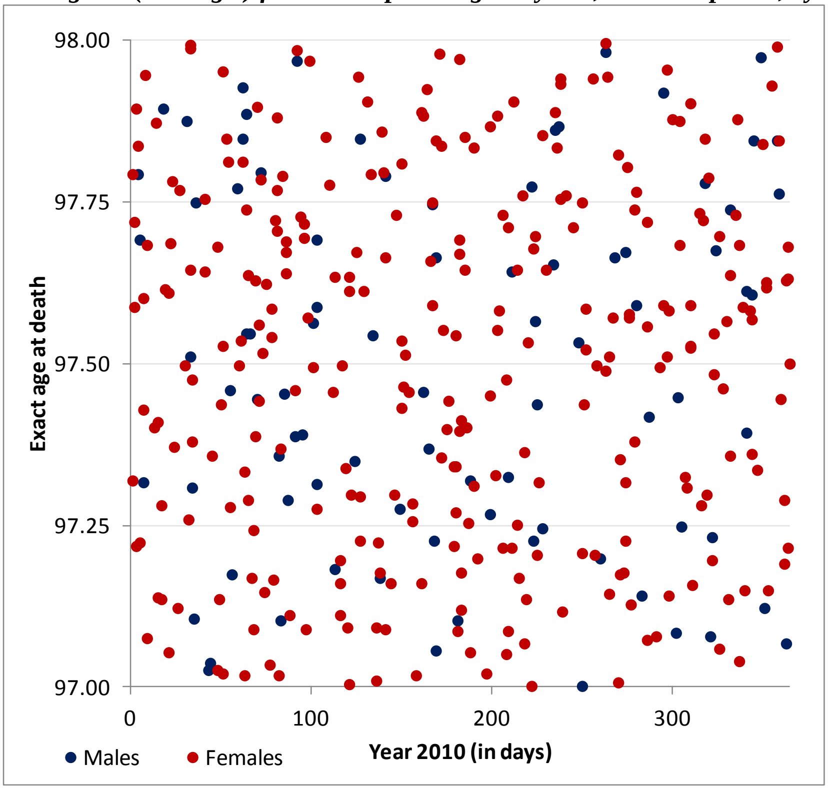
Figure 24: Age-specific mortality rates calculated in different ways, Czech Republic, females, 2010



Note: see Figure 23

Source of data: Author's calculation based on Czech Statistical Office (2011a)

Figure 25: Lexis diagram (rectangle) for the completed age 97 years, Czech Republic, by sex, 2010



Note: Only deaths are shown in the diagram

Source of data: Author's calculation based on Czech Statistical Office (2011a)

Figure 26: Lexis diagram with the distribution of deaths at the completed age 97 years, Czech Republic, males, 2010

Year 2010

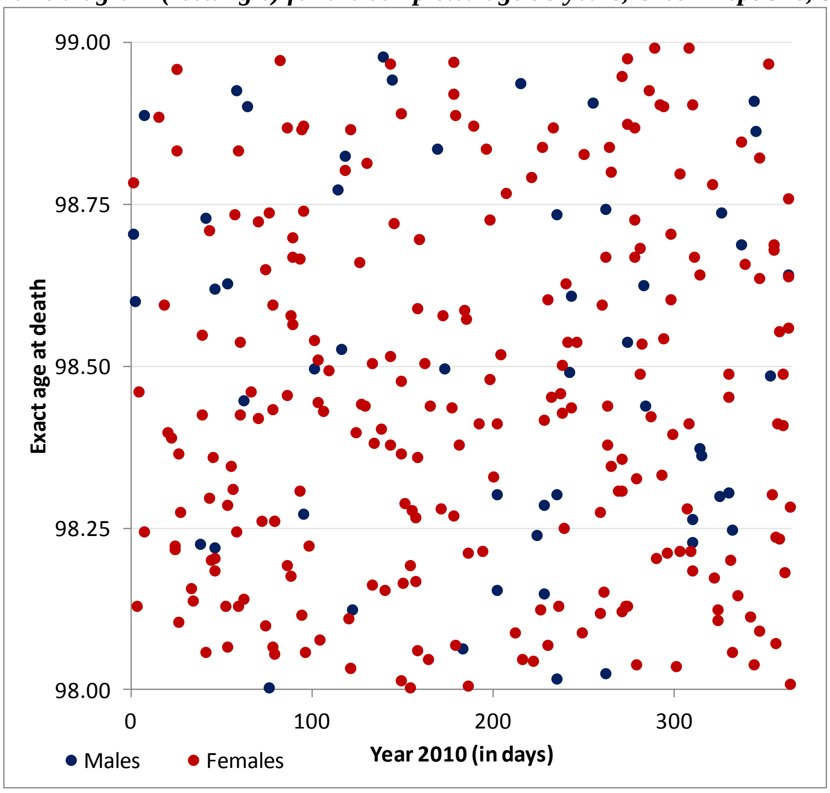
											sum
Age 97		1	1					1	1	1	5
	2	2		1			2			2	9
	2	2		1	1		1		1	2	10
	1		1		1	1	1	2	1	2	10
	1	2	2	1			2	1			9
		2	1		1		1	1	1		7
	2		4	1	1	2				1	11
			1		2	1	2		2		8
	1	1	1	2	1			2		1	9
		2			1		1		2	1	7
sum	9	12	11	6	8	4	10	7	8	10	85

Figure 27: Lexis diagram with the distribution of deaths at the completed age 97 years, Czech Republic, females, 2010

Year 2010

Age 97											sum
	3	1	2	2	3	1	2	3	2	2	21
	3	4	2	1	5	4	3	2	3	3	30
	4	3	6	3	2	3	4	3	3	1	32
	6	4	5	6	3	2	3		2	7	38
	1	4	3		5	2	3	5	7	5	35
	5	3	2	2	4	4	1	2	3	1	27
	3	3	1	1	5	3	1	3	2	3	25
	4	3	1	5	4	4	2	2	2	2	29
	3	2	3	5	1	5	1	4	3	4	31
	2	4	3	5	2	5	1	3	1	1	27
sum	34	31	28	30	34	33	21	27	28	29	295

Figure 28: Lexis diagram (rectangle) for the completed age 98 years, Czech Republic, by sex, 2010



Note: Only deaths are shown in the diagram

Source of data: Author’s calculation based on Czech Statistical Office (2011a)

Figure 29: Lexis diagram with the distribution of deaths at the completed age 98 years, Czech Republic, males, 2010

	Year 2010										sum
Age 98		2		2		1	1			1	7
	1			1	1					1	4
	1	1		1			1	1	1		6
	1	2					1	1		2	7
				1				1			2
		1	1		1		1	1		1	6
						1	1		2	1	5
		2	1				2		3	1	9
				1		1	1				3
			1			1	1	1			4
sum	3	8	3	6	2	4	9	5	6	7	53

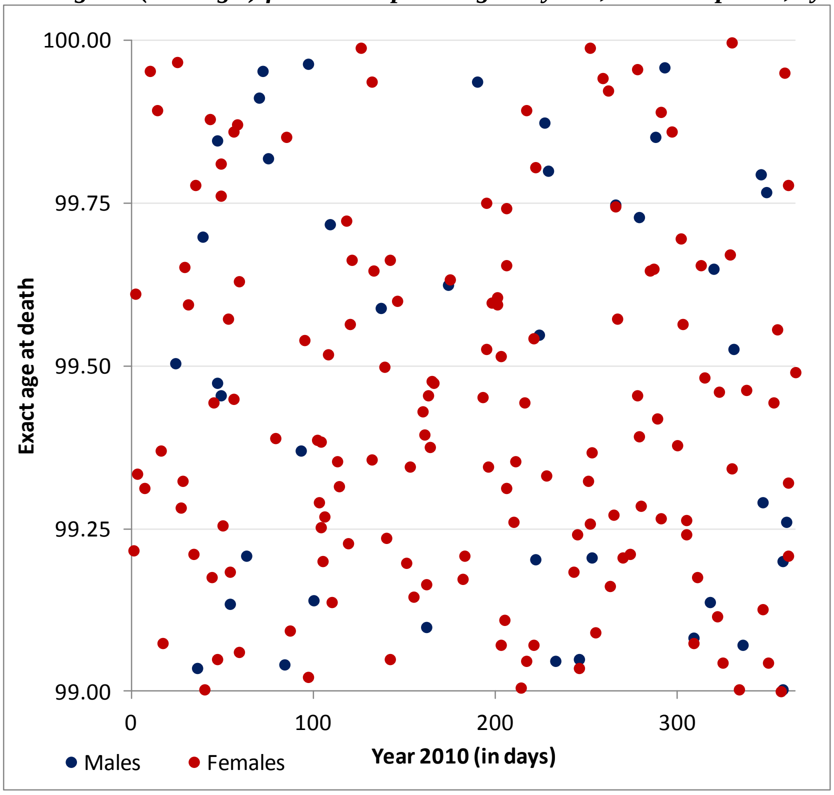
Figure 30: Lexis diagram with the distribution of deaths at the completed age 98 years, Czech Republic, females, 2010

	Year 2010										sum
Age 98	1		1	1	2			5	3	1	14
	2	1	3	3	2	2	3	4		2	22
	1	3	2	1		2	1	1	3	1	15
			4	1	1		2	3	3	5	19
	1	2	5	2	3	3	3	2	1	2	24
	1	4	5	3	3	3	5	3	1	5	33
	3	3	1	3	3	1		6	2	1	23
	4	6	3		5	2	1	2	4	4	31
	4	4	3	3	3		2	5	4	3	31
		2	6	3	5	3	3	1	1	5	29
sum	17	25	33	20	27	16	20	32	22	29	241

All the three pictures (Figures 25, 28 and 31) show the distribution of deaths in the Lexis diagram. With increasing age the number of deaths decreases and also the distribution of deaths is less uniform within the time period and also within the year of age. The uniformity of the distribution of deaths could be evaluated from the graphs where the deaths are summed up and divided into 100 equally long time- and age- intervals (Figures 26, 27, 29, 30, 32 and 33).

According to the results it could be concluded that there are differences caused by the used method of calculation of the age-specific mortality rate. These differences are more significant at higher ages where fewer deaths occur and to which less persons survive. The best and most accurate method of calculation would be that using only the individual lengths of life (Calot, Franco, 2002). In practice, however, such data are usually not achievable and we have to use some compromises. If it is possible to use individual data for people deceased during the studied year their usage seems to be better method than to use only the estimates of the exposure time.

Figure 31: Lexis diagram (rectangle) for the completed age 99 years, Czech Republic, by sex, 2010



Note: Only deaths are shown in the diagram

Source of data: Author's calculation based on Czech Statistical Office (2011a)

It was concluded that the optimal method of calculation of the age-specific mortality rates is the usage of individual life durations during the year. Because age-specific mortality rates are the base for the life table construction it is important to consider these issues before the construction. In the next part of this chapter life tables from the individual data were constructed and the individual life durations were used for the calculation of the exposure time.

Figure 32: Lexis diagram with the distribution of deaths at the completed age 99 years, Czech Republic, males, 2010

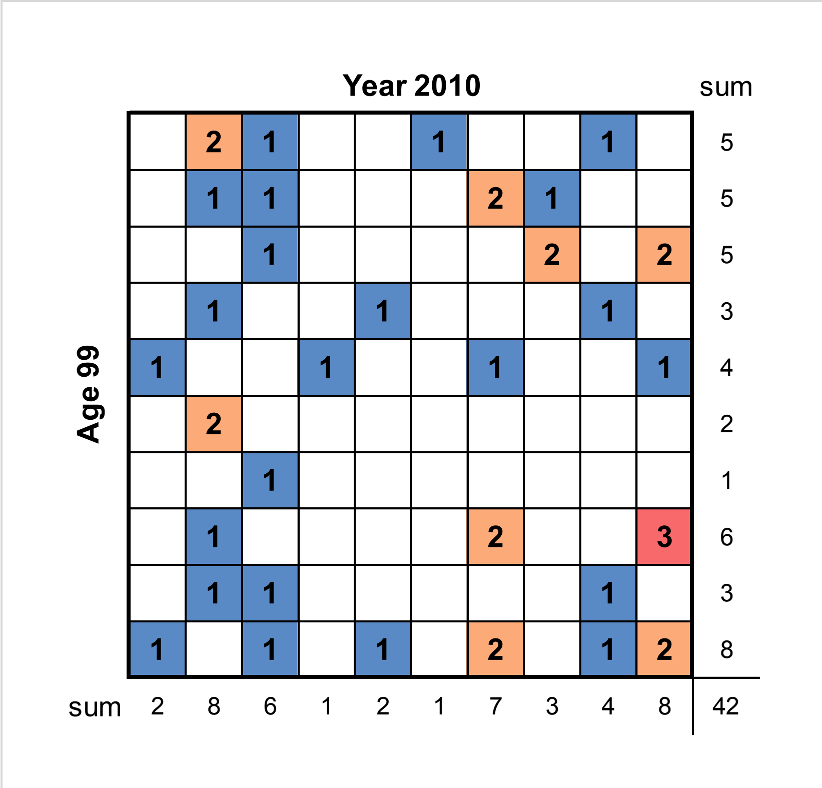
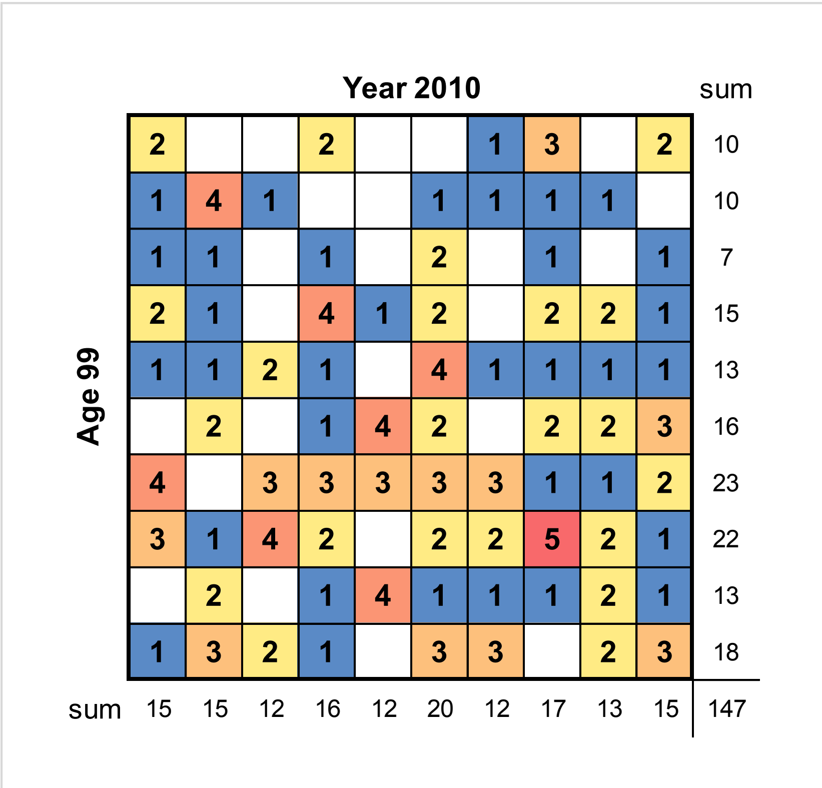


Figure 33: Lexis diagram with the distribution of deaths at the completed age 99 years, Czech Republic, females, 2010



6.2 Life tables of adults of the Czech Republic according to education levels

In the previous part of this chapter several methods of the calculation of the age-specific mortality rates were mentioned. This part could be understood as a practical application of one of them. The method based on individual data about deceased persons was selected. The reason for this choice is simple: it follows the fact that the aim of this part was to construct the life tables not only according to sex but also according to education levels. This combination of characteristics causes that there are relatively rare data in some cases. As a consequence of that the less accurate method could lead to different results. Method of the construction of the life tables according to education will be described in more detail below. For comparison, the life tables according to education levels were constructed also by the traditional method (where the exposure time was estimated as the average of survivors at the beginning and at the end of the particular year). The comparison is done only for tables prepared for data in the 3rd class of deaths and it will be presented in the first part of the rest of this chapter; in the second part the life tables calculated for data in all three classes of deaths will be compared.

6.2.1 Data used for the life tables according to education levels

For the calculation of the life tables according to education levels two main data sources are used. The first of them are the numbers of deaths, individual (anonymous) data collected by the Czech Statistical Office (2011a) are used. For each person deceased during the year 2010 (later, where life tables are calculated from the data in the 1st class of deaths, also data from year 2009 will be used) also the level of education attainment is known together with the date of birth and date of death. So it is possible to calculate the exact age at death and also the life duration during the studied year at the completed age x (rectangles from the Lexis diagram). Then also the individual life durations are calculated during the studied year and particular birth cohort of deceased persons (2nd class of events) and individual life durations according to completed age and cohort of deceased persons (1st class of events).

Except from the data about deaths we need also data about the age- and education-structure of the whole population. Such data are not traditionally published or even collected by the Czech Statistical Office. Therefore it was necessary to use estimates of these structures. In the Czech Republic, there is a possibility to use data from the demographic forecast made according to education levels. This forecast was prepared within the scientific project "RELIK" (Reproduction of the Human Capital) solved by the experts from the University of Economics in Prague and Institute for Information on Education[8]. The forecast was prepared in two different variants, one of them could be understood as a more optimistic one (originally it is labeled as "NL variant" because it is based on the similarity of the development in the Czech Republic with the development in the Netherlands) the other as more pessimistic one (originally labeled as "CZSO variant" because it is based on the previous forecast of the Czech Statistical Office). For the calculations in this Thesis the more optimistic variant was applied. From the forecast the age-structures were taken according to education levels. Education levels are distinguished into

[8] http://www.isvav.cz/projectDetail.do?rowId=2D06026

4 categories: primary education (containing also the category "no education"), lower secondary education, upper secondary education and tertiary education. These categories were used also for construction of the life tables. More details about this forecast could be found in the article by Fiala *et al.* (2011).

In order to have the same categories of education it was necessary to modify the categories of education among deaths. Categories defined as lower secondary education, upper secondary education and tertiary education were the same in both data sets. For the last category all the other levels of education among deaths (primary education, no education and "unknown") were summed up. It must be kept in mind that the category labeled "no and primary education" contains also the cases where the education was unknown and so that the rates calculated for this category may be biased. On the other side there was no other possibility how to solve this problem.

Because the education attainment changes mostly at lower ages, the demographic forecast according to education was calculated only for ages higher than 24 years. In accordance to that also the data about deaths were taken only for ages 25 and higher. As a result also the final life tables were calculated only for ages 25 and higher, so the life expectancy at the first row of these tables is the life expectancy at the exact age 25.

6.2.2 Method of calculation of the age-specific mortality rates

After the data were prepared the age-specific mortality rates had to be calculated. As was stated above, the most accurate method as possible was selected. The calculation was started by the estimation of the exposure time. As was described above, the exposure time was calculated separately for the survivors and for the deceased people from the studied population with distinguishing the elementary classes (triangles) of data.

In the first step the exposure time for the survivors was estimated. For the 3rd class of deaths it was calculated as the sum of the number of persons alive at the beginning of the year and at the end of the year without the number of deaths of those persons who entered the particular Lexis rectangle at the beginning of the year (upper triangle of deaths). This sum was multiplied by 0.5 because all those people spend on average half a year alive in the rectangle defined by the calendar year and completed age (shown above). For life tables constructed for the case of the 2nd class of deaths the exposure time of survivors was estimated from the number of survivors at the end of each particular year because it could be supposed that these persons spent the whole year alive in the studied population (we abstract away from migration). When life tables were constructed for the 1st class of deaths the exposure time of survivors was estimated from the survivors to exact age $x + 1$ (those persons survived the completed age x during calendar years t and $t + 1$). Survivors to exact age $x + 1$ were calculated from the number of survivors at the end of the calendar year t, from this number all relevant deaths were subtracted (the upper triangle of deaths in calendar year $t + 1$). Thanks to this way of calculation the possible effect of migration was eliminated for the 1st class of deaths.

The calculation of the exposure time of deceased persons was more complicated. According to the date of birth and date of death all the people were divided into two groups (two cohorts): in the Lexis diagram they were divided to the upper or lower triangle. For the upper triangle of deaths the individual exposure times were calculated as the difference between the date of death and the beginning of the year, for the lower triangle the individual exposure times were calculated as the difference between the date of death and the date of the x^{th} birthday of the person (i.e. the date when the person entered the studied triangle). All those individual exposure times were then summed up according to the used classification of events (Figure 19) and together with the exposure time of the survivors give the total exposure time, i.e. the denominator of the calculated age-specific mortality rates. In the nominator there was always the number of deaths in the same category. The calculation was repeated for all the categories defined by the education level and sex and for all the three classes of deaths.

From the calculated values of the age-specific mortality rates the life tables were already constructed by the standard method described in more detail e.g. in Calot, Franco (2002). All the tables could be found in the electronic Appendix of this Thesis.

6.2.3 Comparison of two possible methods of calculation

Because the aim of this chapter is not only to construct the life tables according to education attainment but also to illustrate the possible inaccuracy of the life tables constructed by the standard method of time exposure estimation, the tables were constructed also by the standard method. For comparison only the tables constructed for the 3rd class of deaths were used in this sub-chapter. Below the main results will be compared.

For the fulfillment of this goal we have a set of life tables for the Czech Republic constructed by the most detailed method as possible, described above, using the individual life durations. It is then easy to calculate the same set of life tables using the more traditional method of the time exposure estimation – the average of the number of living persons at the beginning and at the end of the year. Then the most important differences were searched. For this purpose the empirical data (rates or probabilities) were not smoothed by any method.

As could be seen from the comparison (see the Table 2) the differences caused by the different methods of time exposure estimation are not significant, especially at lower ages. According to both the methods the life expectancy is the highest for people with the highest level of education. On the other hand significantly lower life expectancy could be seen for males with no or only primary level of education. However, it was mentioned earlier that it has to be kept in mind that the results of this education group could be biased by the influence of persons where the education level was unknown. But it was proven that at higher ages, where fewer deaths occur, the differences between both the used methods are relatively more important, what verifies also the conclusions of Calot and Franco (2002).

For females almost the same as for males could be stated. The differences of results calculated by both the presented methods do not seem to be significant at lower ages. With the increasing age the absolute difference does not change a lot and also the relative difference (in %) grows more slowly than in the case of males. In the case of females the group with the lowest level of education is not as different in its value of the life expectancy as it was in case of males. It is also possible to compare directly the age-specific mortality rates calculated according to the standard method or by the method of individual life durations (Tables 3–8).

Table 2: Comparison of calculated life expectancies according to education attainment, standard method of calculation and method of individual life durations, 3rd class of deaths, selected ages, Czech Republic, both sexes, 2010

Life expectancy (method) / Differences	All education levels	Primary education, no education or unknown	Lower secondary education	Higher secondary education	Tertiary education
MALES					
Age 25 (standard method)	50.14	33.93	53.80	60.12	63.98
Age 25 (indiv. life durations)	50.27	34.08	53.95	60.32	64.21
Difference	0.13	0.15	0.14	0.21	0.22
Relative difference (%)	0.26	0.43	0.27	0.35	0.35
Age 65 (standard method)	15.35	8.02	18.25	23.10	25.63
Age 65 (indiv. life durations)	15.50	8.21	18.41	23.32	25.87
Difference	0.15	0.19	0.16	0.22	0.23
Relative difference (%)	0.96	2.38	0.89	0.97	0.91
Age 80 (standard method)	6.78	3.52	9.33	12.23	13.56
Age 80 (indiv. life durations)	6.96	3.73	9.53	12.50	13.83
Difference	0.18	0.20	0.20	0.27	0.27
Relative difference (%)	2.65	5.76	2.13	2.19	1.96
FEMALES					
Age 25 (standard method)	56.15	49.61	57.56	64.03	67.28
Age 25 (indiv. life durations)	56.28	49.74	57.69	64.18	67.33
Difference	0.13	0.12	0.13	0.15	0.05
Relative difference (%)	0.23	0.25	0.22	0.23	0.08
Age 65 (standard method)	18.80	16.60	20.20	25.40	28.22
Age 65 (indiv. life durations)	18.94	16.75	20.33	25.55	28.27
Difference	0.14	0.14	0.14	0.15	0.06
Relative difference (%)	0.75	0.87	0.67	0.60	0.20
Age 80 (standard method)	8.03	6.92	9.67	13.12	14.92
Age 80 (indiv. life durations)	8.20	7.09	9.82	13.29	14.98
Difference	0.17	0.18	0.16	0.17	0.06
Relative difference (%)	2.11	2.55	1.63	1.29	0.38

Source of data: author's calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

Table 3: Age-specific mortality rates calculated by the standard method of calculation, 3rd class of deaths, ages 80–95 years, Czech Republic, males, 2010

Age	All education levels	Primary education, no education or unknown	Lower secondary education	Higher secondary education	Tertiary education
80	0.085020	0.228965	0.057571	0.032186	0.020926
81	0.095101	0.248225	0.064834	0.035939	0.022131
82	0.105588	0.259633	0.071467	0.039805	0.034722
83	0.120743	0.290460	0.078811	0.045748	0.043889
84	0.129602	0.309789	0.085657	0.045301	0.043134
85	0.142494	0.355748	0.087722	0.044307	0.046693
86	0.164570	0.398131	0.101085	0.064851	0.059477
87	0.177567	0.405781	0.121212	0.060150	0.062805
88	0.198897	0.440762	0.141456	0.072314	0.060161
89	0.236595	0.552533	0.139167	0.078890	0.097462
90	0.258587	0.591292	0.159151	0.108671	0.069731
91	0.245712	0.579560	0.142857	0.077922	0.103448
92	0.266496	0.604651	0.168539	0.066667	0.111732
93	0.318105	0.746269	0.151079	0.159509	0.125000
94	0.326198	0.691030	0.151844	0.222222	0.156863
95	0.382418	0.771812	0.170213	0.244898	0.241758

Source of data: author’s calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

Table 4: Age-specific mortality rates calculated by the method of individual life durations, 3rd class of deaths, ages 80–95 years, Czech Republic, males, 2010

Age	All education levels	Primary education, no education or unknown	Lower secondary education	Higher secondary education	Tertiary education
80	0.083157	0.216176	0.056683	0.031904	0.020815
81	0.092859	0.232949	0.063855	0.035604	0.022028
82	0.102791	0.243259	0.070207	0.039383	0.034408
83	0.117113	0.270438	0.077247	0.045228	0.043350
84	0.125561	0.287855	0.083899	0.044722	0.042673
85	0.137500	0.325576	0.085860	0.043817	0.046207
86	0.158270	0.363243	0.098723	0.063803	0.058579
87	0.170221	0.369440	0.117875	0.059211	0.061723
88	0.189679	0.397087	0.137045	0.070913	0.059295
89	0.222811	0.482253	0.134276	0.076943	0.095912
90	0.244168	0.523516	0.153043	0.105998	0.068687
91	0.232746	0.512207	0.138482	0.076577	0.100798
92	0.251814	0.538034	0.162399	0.065511	0.107749
93	0.295744	0.627825	0.146543	0.153441	0.122594
94	0.304954	0.603867	0.147542	0.205376	0.154061
95	0.346927	0.642261	0.160417	0.230848	0.238640

Source of data: author’s calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

It is clear, that in the case of males the age-specific mortality rates calculated by the traditional method are a bit higher in comparison to the values calculated by the method of individual life durations; this difference increases with age. At higher ages the relative differences reach values up to almost 10 % (for the group of the lowest level of education even higher, see Table 5).

Table 5: Relative difference of age-specific mortality rates calculated by the standard method and by the method of individual life durations, 3rd class of deaths, ages 80–95 years, Czech Republic, males, 2010 (in %)

Age	All education levels	Primary education, no education or unknown	Lower secondary education	Higher secondary education	Tertiary education
80	–2.2	–5.6	–1.5	–0.9	–0.5
81	–2.4	–6.2	–1.5	–0.9	–0.5
82	–2.7	–6.3	–1.8	–1.1	–0.9
83	–3.0	–6.9	–2.0	–1.1	–1.2
84	–3.1	–7.1	–2.1	–1.3	–1.1
85	–3.5	–8.5	–2.1	–1.1	–1.0
86	–3.8	–8.8	–2.3	–1.6	–1.5
87	–4.1	–9.0	–2.8	–1.6	–1.7
88	–4.6	–9.9	–3.1	–1.9	–1.4
89	–5.8	–12.7	–3.5	–2.5	–1.6
90	–5.6	–11.5	–3.8	–2.5	–1.5
91	–5.3	–11.6	–3.1	–1.7	–2.6
92	–5.5	–11.0	–3.6	–1.7	–3.6
93	–7.0	–15.9	–3.0	–3.8	–1.9
94	–6.5	–12.6	–2.8	–7.6	–1.8
95	–9.3	–16.8	–5.8	–5.7	–1.3

Note: negative values signify that the age-specific mortality rates calculated by the method of individual life durations are lower than the rates calculated by the standard method

Source of data: author’s calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

Table 6: Age-specific mortality rates calculated by the standard method of calculation, 3rd class of deaths, ages 80–95 years, Czech Republic, females, 2010

Age	All education levels	Primary education, no education or unknown	Lower secondary education	Higher secondary education	Tertiary education
80	0.056096	0.073065	0.043334	0.026472	0.011692
81	0.064733	0.088215	0.042178	0.028371	0.012479
82	0.072905	0.095377	0.050435	0.035748	0.015399
83	0.087911	0.112047	0.066250	0.034627	0.018549
84	0.099052	0.128435	0.068145	0.036492	0.025078
85	0.110752	0.141767	0.078804	0.039716	0.028853
86	0.131177	0.166373	0.098330	0.043358	0.033778
87	0.141719	0.177450	0.111089	0.049190	0.022222
88	0.166871	0.213721	0.112925	0.065225	0.069465
89	0.178064	0.221335	0.130343	0.084630	0.023715
90	0.219605	0.279764	0.152236	0.078148	0.057895
91	0.215626	0.271046	0.157419	0.082051	0.034335
92	0.250363	0.313269	0.171115	0.099010	0.045455
93	0.247555	0.300358	0.177370	0.097902	0.039216
94	0.289997	0.342660	0.223030	0.103004	0.022727
95	0.343759	0.392919	0.259603	0.201923	0.076923

Source of data: author’s calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

Again it was shown that in the case of females the differences are lower in comparison to males. For both sexes at lower ages the differences are significantly lower in comparison to higher ages. It also confirms the results of Calot and Franco (2002) mentioned above. It could be concluded that the method of construction of the exposure time matters above all at higher ages. Both the compared methods reflected the overall trend in the data, which means the differences among various groups defined by the level of education or sex.

Table 7: Age-specific mortality rates calculated by the method of individual life durations, 3rd class of deaths, ages 80–95 years, Czech Republic, females, 2010

Age	All education levels	Primary education, no education or unknown	Lower secondary education	Higher secondary education	Tertiary education
80	0.055286	0.071721	0.042844	0.026259	0.011644
81	0.063672	0.086263	0.041704	0.028203	0.012448
82	0.071553	0.093049	0.049829	0.035364	0.015345
83	0.085946	0.108862	0.065181	0.034228	0.018493
84	0.096623	0.124358	0.067025	0.036130	0.024922
85	0.107661	0.136691	0.077286	0.039303	0.028661
86	0.127009	0.159667	0.096041	0.042895	0.033488
87	0.136996	0.170092	0.108297	0.048513	0.022010
88	0.160260	0.202896	0.109984	0.064153	0.068370
89	0.170321	0.209258	0.126408	0.082904	0.023614
90	0.209050	0.262950	0.146949	0.076714	0.057391
91	0.204854	0.254067	0.151753	0.080703	0.033862
92	0.236453	0.291021	0.165572	0.096737	0.044987
93	0.232309	0.278115	0.170339	0.094091	0.038586
94	0.270350	0.315501	0.211180	0.100261	0.022870
95	0.320409	0.361023	0.248230	0.198447	0.075196

Source of data: author's calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

Table 8: Relative difference of age-specific mortality rates calculated by the standard method and by the method of individual life durations, 3rd class of deaths, ages 80–95 years, Czech Republic, females, 2010 (in %)

Age	All education levels	Primary education, no education or unknown	Lower secondary education	Higher secondary education	Tertiary education
80	–1.4	–1.8	–1.1	–0.8	–0.4
81	–1.6	–2.2	–1.1	–0.6	–0.3
82	–1.9	–2.4	–1.2	–1.1	–0.4
83	–2.2	–2.8	–1.6	–1.2	–0.3
84	–2.5	–3.2	–1.6	–1.0	–0.6
85	–2.8	–3.6	–1.9	–1.0	–0.7
86	–3.2	–4.0	–2.3	–1.1	–0.9
87	–3.3	–4.2	–2.5	–1.4	–1.0
88	–4.0	–5.1	–2.6	–1.6	–1.6
89	–4.4	–5.5	–3.0	–2.0	–0.4
90	–4.8	–6.0	–3.5	–1.8	–0.9
91	–5.0	–6.3	–3.6	–1.6	–1.4
92	–5.6	–7.1	–3.2	–2.3	–1.0
93	–6.2	–7.4	–4.0	–3.9	–1.6
94	–6.8	–7.9	–5.3	–2.7	0.6
95	–6.8	–8.1	–4.9	–1.7	–2.3

Note: negative values signify that the age-specific mortality rates calculated by the method of individual life durations are lower than the rates calculated by the standard method

Source of data: author's calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

In accordance to the results of this comparison the theoretically more accurate method, using the individual life durations, was used for the construction of the life tables according to education attainment for both sexes. In the next sub-chapter the life tables using data classified into the 1st, 2nd and also the 3rd class of events were calculated so as it will be possible to see the differences and to use the full advantage of the individual data.

6.2.4 Main results and basic comparison of life tables constructed for the 1st, 2nd, and the 3rd class of deaths and according to education levels

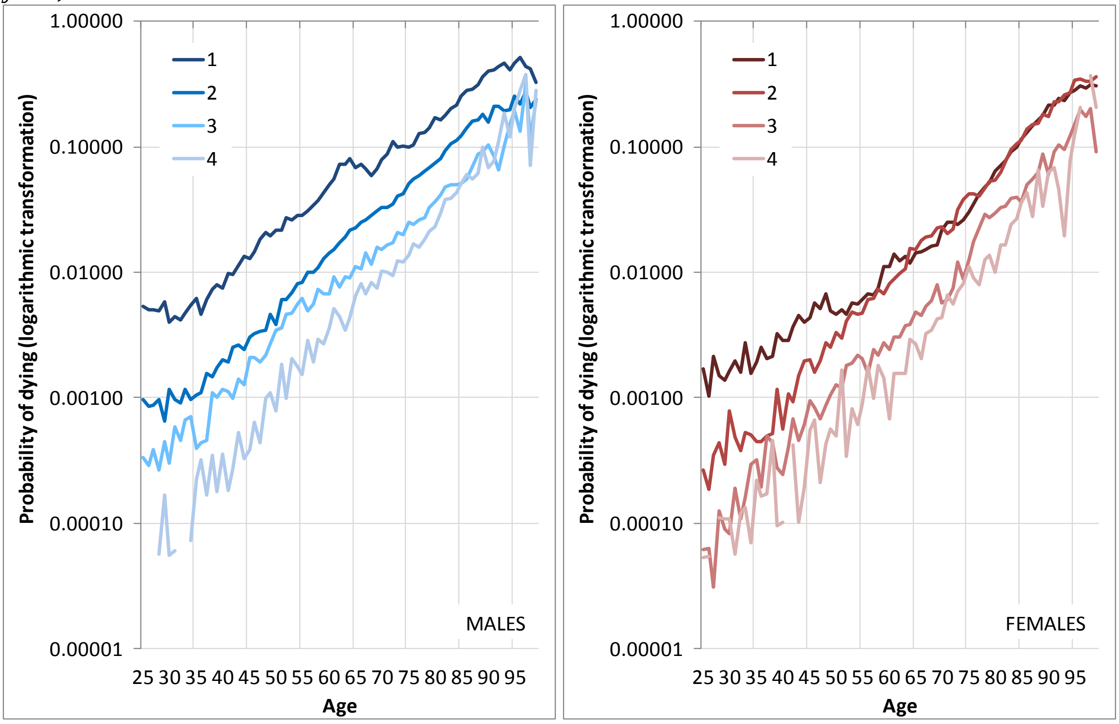
Although all the tables are attached to this Thesis it the Appendix, in this chapter some main results will be presented. The aim of the analysis was to calculate the life tables according to sex and education attainment for the Czech Republic. Because the individual data about deceased persons were available it was possible to reconstruct the classification to elementary classes (triangles) of the Lexis diagram. These triangles were used for the construction of tables according to 1st, 2nd, and 3rd class of deaths. The first mentioned type of tables enables to study the process of mortality according to the completed age and also the original cohort of people during two calendar years (in our case these two years are 2009 and 2010). The second type enables to characterize mortality for particular cohorts during the studied year (2010). The 3rd type, the most traditional way of construction of life tables, shows the mortality situation according to completed age during a studied year (2010) but two different cohorts of persons are in that case mixed together.

Later in the text there is no need to describe the method of calculation because it was fully described above at the beginning of this chapter. Generally in all cases the data about deaths were classified into the elementary classes (triangles) of the Lexis diagram. For both types of triangles (upper and lower) the proper methodology was used for the calculation of individual life durations of deceased and also of surviving people. These life durations were summed up and formed the denominator of the age-specific mortality rates. In the nominator there was always the number of deaths classified in the same manner. The age-specific mortality rates were the input for the life table. Then the standard method of life table construction was used. All the tables were ended by the last open age-interval 100+ years (in the case of tables constructed for the 2nd class of deaths the last row was calculated as the open interval of cohorts – 1909 and older). For all those open intervals the probability of dying was set to be equal to one.

The life table calculated for the 2nd class of deaths has one specific feature more – on the first row there the data for only one triangle (characterized by the completed age x = 25, year 2010, cohort $2010 - x = 2010 - 25 = 1985$) are contained. Only on the second row there start the data traditionally classified into the 2nd class, on the second row there are data about mortality at completed ages 25–26, cohort 1984.

In the Figure 34 it is possible to see the basic results calculated for data classified into the 1st classes (the arrangement of the classes of data used for the calculation of the life tables could be seen in the picture in Figure 35). For comparison of the mortality pattern the function of probability of dying was selected. For males so as for females the differences according to education attainment were confirmed. As could be seen there are important differences for males according to education level – only for the oldest-old it seems that the probability of dying for the two highest education levels (higher secondary and tertiary education) is almost the same. In all ages there is significant difference between males with no or only primary education and the other education levels. This difference persists in all ages.

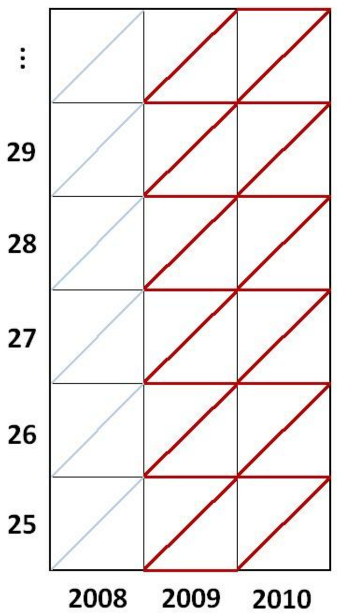
Figure 34: Probability of dying according to education attainment, 1st class of deaths, Czech Republic, by sex, 2009–2010



Note: In the legend of the graph, there are the particular levels of the education. "1" stands for the category "primary education, no education or unknown", "2" stands for the lower secondary education, "3" stands for the higher secondary education and "4" for the tertiary education.

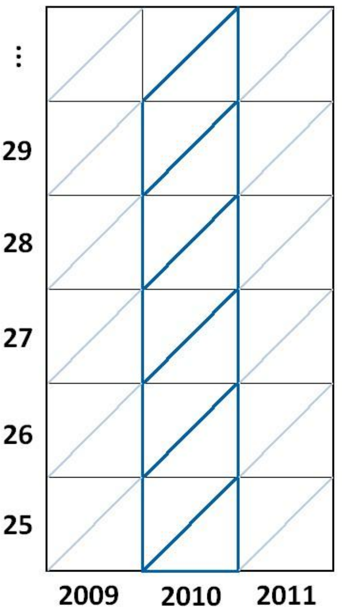
Source of data: author's calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

Figure 35: 1st classes of deaths used for the life table construction



It is possible to see an interesting fact for females with no or primary education and with lower secondary education. For ages above ca 50 years the probability of dying of those two groups of persons are almost identical, in some ages even the probability of dying is lower for females with no or primary education in comparison with the other group. At younger ages this similarity diminishes and the difference between those two groups of females increases. This result is in harmony with the results of Rychtaříková (2010). On the other hand the difference between females with higher secondary education and tertiary education is not as visible as in the case of males at lower ages. One possible reason for the results could be the life style which is different for both sexes – while females do not differ so much according to education attainment it could be supposed that the life style of males is more influenced by education (it has the impact to the type of work, physical activity or psychical condition). All those factors are suggested also by Rychtaříková (2010) who studies also the differences according to various causes of death.

Figure 36: 2nd classes of deaths used for the life table construction



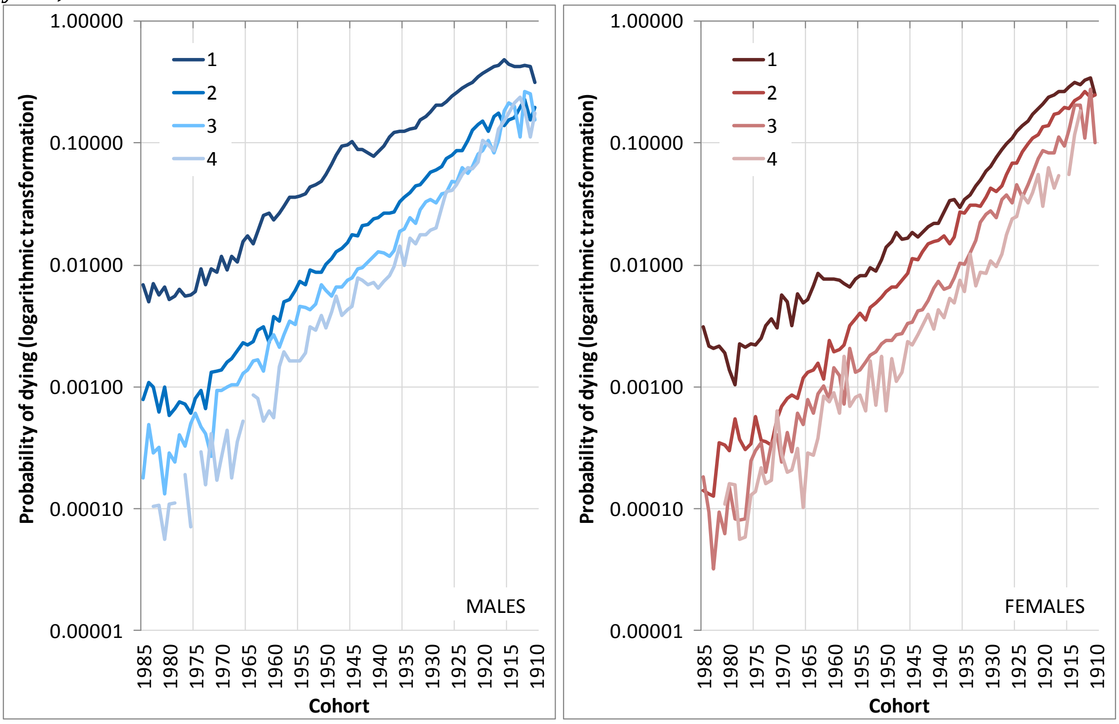
In the Figure 36 there the arrangement of the 2nd classes of deaths could be seen as it was used for the calculation of life tables. As was mentioned above, there is one special feature in the data. At the beginning of the life table there on the first row the data are relevant not for the whole class but only for the lower triangle (for completed age 25, calendar year 2010 and cohort *2010 – 25 = 1985*). From that row up already at each row the data corresponds with one particular cohort in the year 2010 (with two completed ages). In the first column of such a table there are not the values of age but the years characterizing particular cohorts. The youngest cohort in our life table is the cohort 1985. On the last row of the table there is also the open cohort-interval – the data represents people born in 1909 or before that year.

Again for the comparison of mortality conditions according to age we can use the function of the probability of dying. From the graphs (Figure 37) it could be seen that the general pattern of mortality according to age and education attainment is nearly the same as was shown for the 1st class of deaths. But it is clear that the differences among various education levels are bigger, especially the group with the lowest level of education seems to be more different than in the previous case (1st classes).

It is very difficult to find any possible reason for this fact. One of them may be the difference in the composition of the 1st and 2nd classes of demographic events. The 1st classes respect the age- and cohort-specificity, on the other hand the 2nd classes contains only one generation but two completed ages which are mixed together. In accordance to that theoretically the bigger differences in the results when the 2nd classes of deaths were used could be the result of the mixture of two different ages in one class. It would mean that the differences between ages are more significant than differences between calendar years (because the 1st classes are mixtures of data from two calendar years).

It must be also added that the 1st class of deaths is the only one which is thanks to the calculation of the mortality rates (described above) fully purified from the effect of migration. As would be shown later, the similarity of the results for the 2nd and 3rd classes of deaths could be taken as a confirmation of these assumptions.

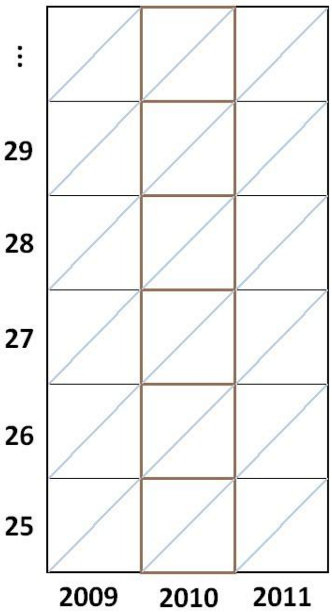
Figure 37: Probability of dying according to education attainment, 2nd class of deaths, Czech Republic, by sex, 2010



Note: In the legend of the graph, there are the particular levels of the education. "1" stands for the category "primary education, no education or unknown", "2" stands for the lower secondary education, "3" stands for the higher secondary education and "4" for the tertiary education.

Source of data: author's calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

Figure 38: 3rd classes of deaths used for the life table construction

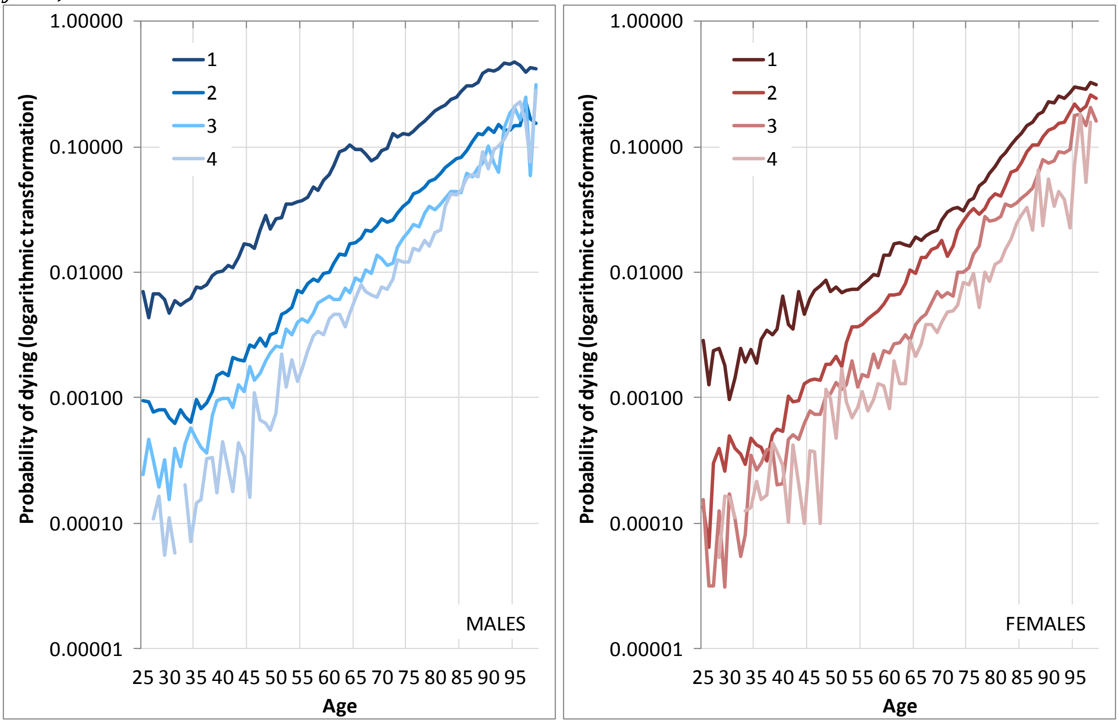


Finally the life tables were constructed for the 3rd classes of deaths. It is probably the most common way of life table construction, its features and method of construction was described in more detail above. The life tables are constructed for the calendar year 2010 so as it could be compared with the previous results obtained for the 1st and 2nd classes of data. The arrangement of the data could be seen in the Figure 38.

Because of the previous results obtained for the 1st and 2nd classes of deaths it could be expected that for the 3rd class the differences among various education levels should be bigger in comparison to the case of the 1st class because the 3rd class is the mixture of two different generations and moreover also in the 3rd class of events the undesirable effect of migration have to be expected.

The results presented in the Figure 39 confirmed the expectations. Again the differences mainly between the lowest educational group and the others are slightly bigger than in the case of the data arranged to the 1st class of events.

Figure 39: Probability of dying according to education attainment, 3rd class of deaths, Czech Republic, by sex, 2010



Note: In the legend of the graph, there are the particular levels of the education. “1” stands for the category “primary education, no education or unknown”, “2” stands for the lower secondary education, “3” stands for the higher secondary education and “4” for the tertiary education.

Source of data: author’s calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

Table 9: Life expectancy at the age 25 according to education attainment, various classifications of input data, Czech Republic, males, 2010

Classification of data / Differences	All education levels	Primary education, no education or unknown	Lower secondary education	Higher secondary education	Tertiary education
1st class of events	50.03	36.78	50.77	58.54	62.63
2nd class of events	50.61	34.49	54.09	60.05	63.42
3rd class of events	50.27	34.08	53.95	60.32	64.21
max – min	0.58	2.70	3.32	1.79	1.58
Official life tables[9]	50.07	–	–	–	–

Source of data: author’s calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

For all the results presented in this part of the chapter it could be stated that the general pattern of mortality according to education attainment remains the same for all the three possible types of data classification. In all three cases it was confirmed that there are significant differences in mortality of the lowest education level and the other groups of education attainment. However, this was confirmed more visibly for males; in case of females it was confirmed only at lower ages. At higher ages the mortality level of females with no or primary education and females with lower secondary education is almost the same. The possible reason for this fact could be the different life style of males and females which is in case of males (and

[9] http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky_za_cr_od_roku_1920/$File/cr_ut_1920_2010.zip

probably younger females) more influenced by the level of education. In all three cases it was confirmed that there is almost no difference in the probability of dying for males with higher secondary education and tertiary education at the highest age groups. It should be only reminded that for the calculation in all the cases the method of individual life duration was used.

Table 10: Life expectancy at the age 25 according to education attainment, various classifications of input data, Czech Republic, females, 2010

Classification of data / Differences	All education levels	Primary education, no education or unknown	Lower secondary education	Higher secondary education	Tertiary education
1st class of events	56.03	52.24	54.24	63.42	66.22
2nd class of events	56.63	50.06	57.96	64.05	67.72
3rd class of events	56.28	49.74	57.69	64.18	67.33
max – min	0.60	2.50	3.72	0.75	1.49
Official life tables[10]	56.07	–	–	–	–

Source of data: author's calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

6.2.5 Smoothed life tables constructed for the 1st, 2nd, and the 3rd class of deaths and according to education levels

From the text and figures above it was clear that the age-specific mortality rates as well as the probabilities of dying were calculated in some cases from quite rare data. It caused the fluctuations visible in the Figures 34, 37 and 39. So as to find the general trend in the data and so as to reveal the more general age-pattern the empirical data were smoothed. The age-specific mortality rates of all combinations of sex and education level were smoothed by the Thatcher function applied to ages 25–85 years. The older ages were extrapolated using the estimations of the parameters of the function. For this purpose the attached macro was used (introduced in the Chapter 5). Then the recalculation for any other choice of the mortality law (function of smoothing and extrapolation) could be very easy and quick. In the macro, there are except for the Thatcher function also the Kannisto, Coale-Kisker, Gompertz, Gompertz-Makeham, and the modified Gompertz-Makeham functions implemented (see Chapter 5 for more details).

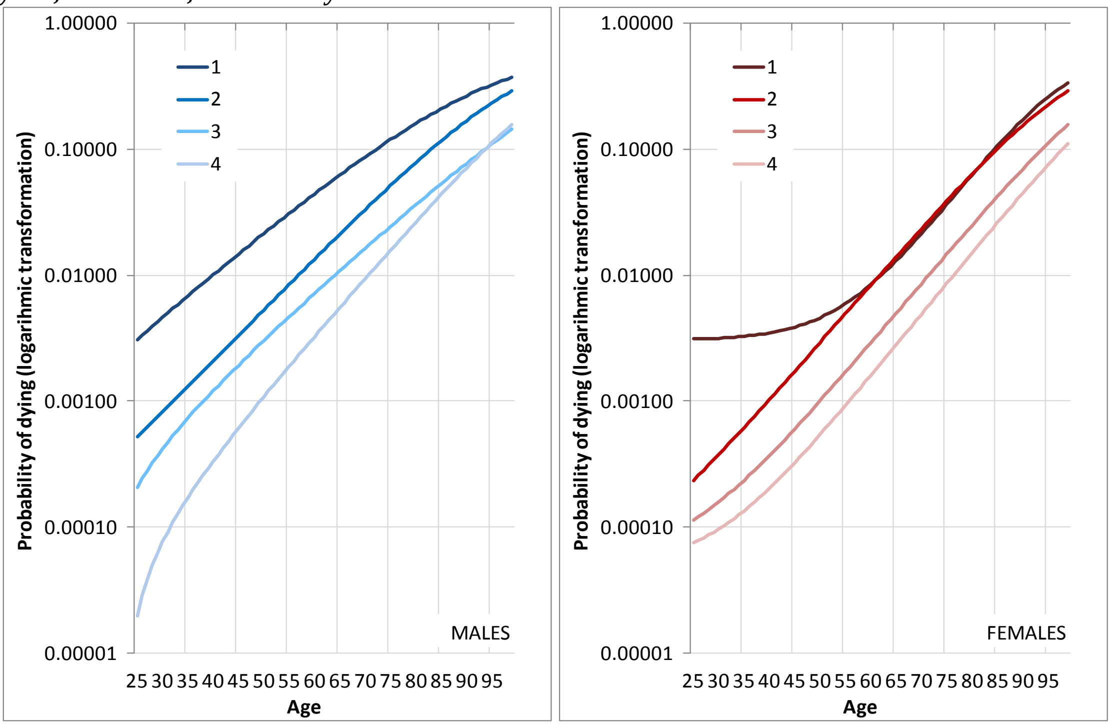
In the SAS macro the age-specific mortality rates were extrapolated up to the age 120. When the empirical rates were smoothed and extrapolated then the traditional life tables were constructed again. So as to end the life tables in the same way as in the previous part, the last row was selected as the open age-interval 100 and more years. For the 2nd class of deaths the last row contains the open cohort-interval 1909 and older. Then the value of the probability of dying at the last row was taken as equal to one. The life table function L_x at the last row was then calculated as a sum of this function for ages 101 and more. Other functions were calculated in the traditional manner.

One more note has to be added for the case of the 2nd class of events: the first row (containing data only for the first triangle) in the unsmoothed tables was not used for the

[10] http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky_za_cr_od_roku_1920/$File/cr_ut_1920_2010.zip

smoothing and it is omitted also in the smoothed tables used in this part. Therefore the first row in the smoothed tables constructed for the 2nd class of deaths contains data for the first whole parallelogram (representing the age 25–26 in the year 2010 and cohort 1984).

Figure 40: Probability of dying according to education attainment, 1st class of deaths, Czech Republic, by sex, 2009–2010, smoothed by the Thatcher method



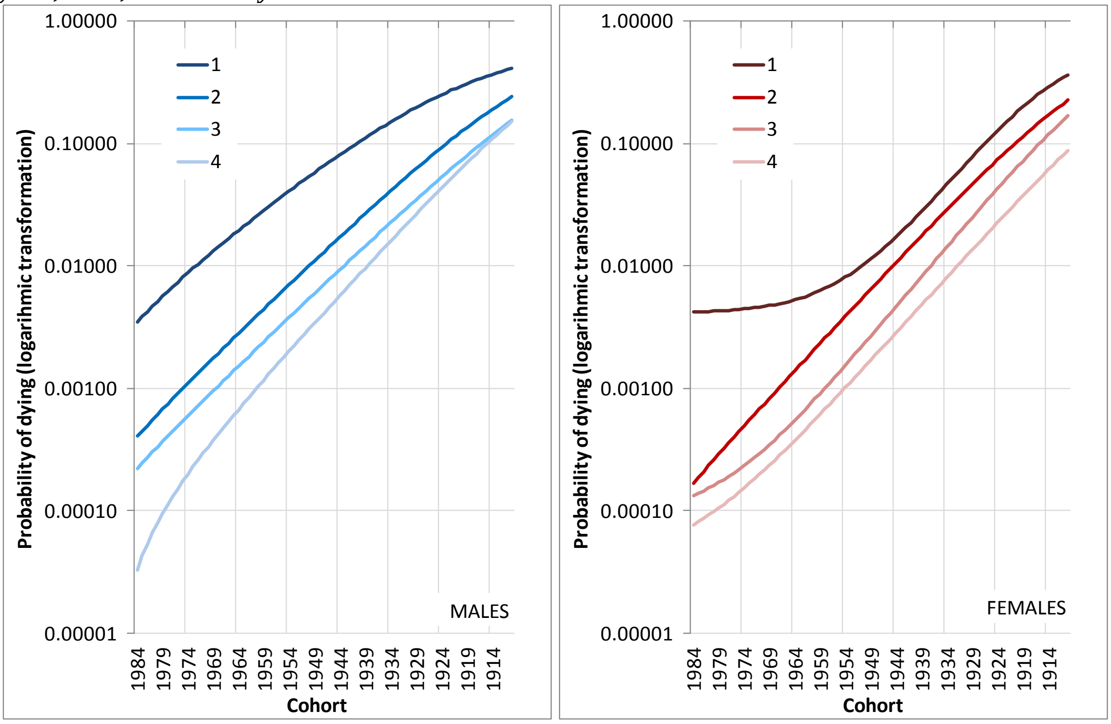
Note: In the legend of the graph, there are the particular levels of the education. “1” stands for the category “primary education, no education or unknown”, “2” stands for the lower secondary education, “3” stands for the higher secondary education and “4” for the tertiary education.

Source of data: author’s calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

As could be seen in the figures, the smoothing method helped to express the trend of the data and its general pattern as was the initial aim. Again the tables were constructed for the data classified into all the three possible classes. For the 1st class of deaths the results could be seen in the Figure 40. The application of the method of smoothing confirmed the facts mentioned already above – the significant differences among various education levels of people. Mainly for males the differences are visible clearly. The biggest difference could be seen between the two lowest levels of education.

For females the interesting anomaly mentioned by Rychtaříková (2010) and confirmed in the previous sub-chapter could be seen again. From the graph it could be stated that in general at higher ages there is nearly no difference of mortality between those groups of females with primary or no education and females with lower secondary education. On the other side the difference between these two groups is significant at lower ages what could represent the changes of the impact of a low level of education to the female’s lifestyle, health and finally also to the risk of death.

Figure 41: Probability of dying according to education attainment, 2nd class of deaths, Czech Republic, by sex, 2010, smoothed by the Thatcher method



Note: In the legend of the graph, there are the particular levels of the education. "1" stands for the category "primary education, no education or unknown", "2" stands for the lower secondary education, "3" stands for the higher secondary education and "4" for the tertiary education.

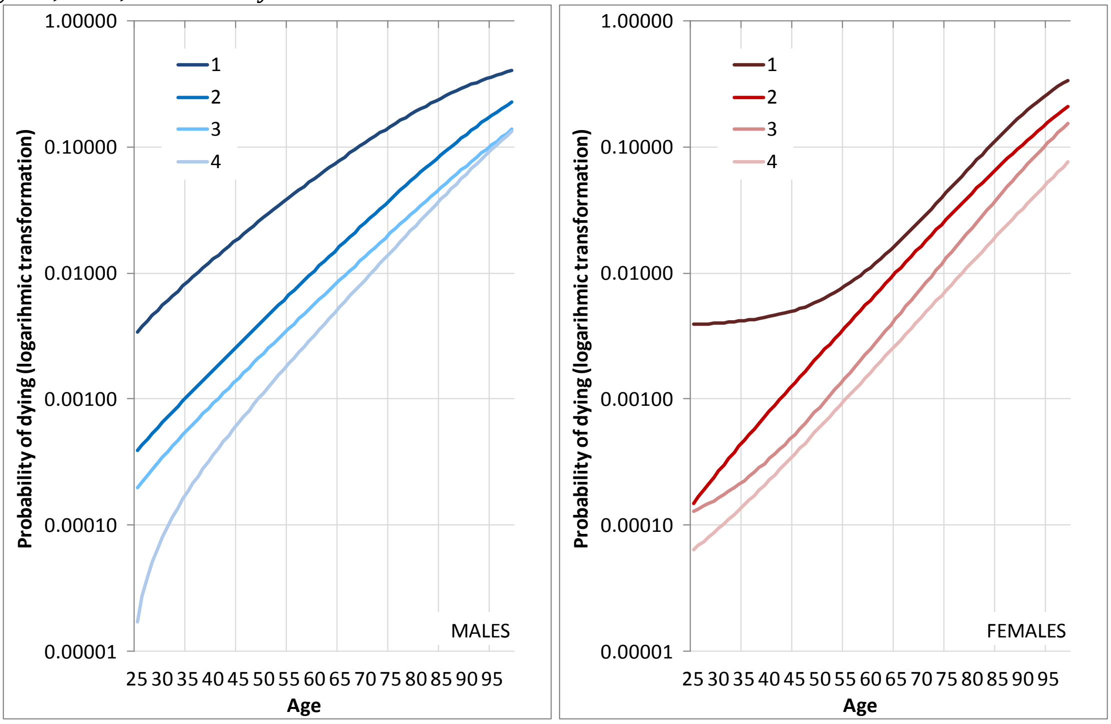
Source of data: author's calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

For the 2nd and 3rd classes of deaths almost the same as in the previous case could be said. Again through the applied method of smoothing the general pattern of the data could be seen. Also in this case the results for the 2nd and 3rd classes of deaths are slightly different from the results obtained for the 1st class of deaths. It is more visible for the case of females, where for the 2nd and 3rd class of deaths a difference remains between the probability of dying of females with primary or no education and females with lower secondary education (see Figures 41 and 42). But still there the probabilities at higher ages lie close to each other.

Also the fact of huge differences among males with different levels of education was confirmed – the most significant difference is again between males with no or primary education and males with lower secondary education. It could be again interpreted as an important impact of the low level of education to the life style of males, to their work, activities, health and finally the risk of death (Rychtaříková, 2010).

Based on the results (summarized in the Figures 40, 41, 42 and Tables 11 and 12) it could be concluded that the selected way of calculation of the life table according to sex and level of education can significantly influence the results. For the purpose of expression the mortality pattern according to educational levels the most accurate method as possible was used (method of individual life durations). The life tables were then calculated for all the three possible classes of deaths and for the purpose of this sub-chapter they were also smoothed and then extrapolated to the highest ages so as the random fluctuations could be eliminated. The results for all three classes are very similar and reveal the same trend and tendencies. However, for the 1st class of deaths the results differ slightly more. It could be, among others, the consequence of the possible impact of migration in the 2nd and 3rd class of demographic events which is possible to eliminate fully only in the case of the 1st class of events.

Figure 42: Probability of dying according to education attainment, 3rd class of deaths, Czech Republic, by sex, 2010, smoothed by the Thatcher method



Note: In the legend of the graph, there are the particular levels of the education. "1" stands for the category "primary education, no education or unknown", "2" stands for the lower secondary education, "3" stands for the higher secondary education and "4" for the tertiary education.

Source of data: author's calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

Table 11: Life expectancy at the age 25 according to education attainment, various classifications of input data, Czech Republic, males, 2010, smoothed by the Thatcher method

Classification of data / Differences	All education levels	Primary education, no education or unknown	Lower secondary education	Higher secondary education	Tertiary education
1st class of events	50.15	36.86	50.77	59.01	63.47
2nd class of events	49.71	33.69	53.15	59.90	63.48
3rd class of events	50.33	34.02	53.80	60.87	64.47
max – min	0.62	3.18	3.02	1.86	1.00
Official life tables[11]	50.07	–	–	–	–

Source of data: author's calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

[11] http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky_za_cr_od_roku_1920/$File/cr_ut_1920_2010.zip

Table 12: Life expectancy at the age 25 according to education attainment, various classifications of input data, Czech Republic, females, 2010, smoothed by the Thatcher method

Classification of data / Differences	All education levels	Primary education, no education or unknown	Lower secondary education	Higher secondary education	Tertiary education
1st class of events	56.28	51.88	54.60	63.92	68.51
2nd class of events	55.87	48.61	57.54	63.94	69.86
3rd class of events	56.52	49.29	58.23	64.67	71.07
max – min	0.65	3.27	3.63	0.75	2.57
Official life tables[12]	56.07	–	–	–	–

Source of data: author's calculation based on Czech Statistical Office (2011a) and Hulík, Fiala (2011)

6.3 Conclusions

In this chapter some basic but important facts related to the issue of life tables were mentioned. On the example it was shown how it is possible to calculate the age-specific mortality rates which are the base for the life table construction. It was shown that it should be preferred to use as detailed data as possible. Ideally the individual data for the calculation of individual exposure times should be used. At lower ages the differences according to various methods of age-specific mortality rates are not significant. At higher ages the differences are more visible.

In the second part of this chapter the most detailed data as possible were used for the construction of life tables calculated for the Czech Republic. Thanks to the accessibility of individual data the life tables according to education levels could be constructed. At first only the 3rd classes of deaths were considered and the calculation was made by the standard method and by the method of individual life durations. The standard method means the estimation of exposure time as an average of the numbers of survivors at the beginning and at the end of the year. Again it was concluded that the results differ above all at higher ages.

Then for the construction of life tables according to education attainment three different methods were used. These methods differ only in the form of data classification. So the data in the 1st, 2nd, and the 3rd classes of demographic events (deaths) were used. In the three sets of constructed life tables it was clearly confirmed the general pattern of mortality according to age and education level. For males there is the most important difference – between males with no or only primary education and other educational levels. At younger ages the same difference was confirmed also for females. However females with no or primary education aged ca 50 or more years have nearly comparable probability of dying as females with lower secondary education. At some ages even the probability of dying was lower for the least educated group of females. For males there is another specific feature according to mortality and education – it is the fact that at the highest ages the differences between males with higher secondary education

[12] http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky_za_cr_od_roku_1920/$File/cr_ut_1920_2010.zip

and tertiary education diminishes. For females there is the difference between these two educational groups only small in the whole age-range.

In the last part of the chapter there the life tables according to education attainment were smoothed. The SAS macro introduced in the previous chapter (Chapter 5) was used for this purpose and the Thatcher logistic method of smoothing was selected. The unknown parameters were calculated from ages 25–85 years, so as the differences also at lower ages could be preserved. At higher ages the selected method was used also for the extrapolation of the intensity of mortality. Then the smoothed life tables were compared and it was stated that the smoothing applied to the data enabled to see clearly the general pattern without any random fluctuations. In this way the results from the previous part were confirmed. All the tables according to education attainment (unsmoothed as well as the smoothed ones) are attached in an electronic form in the Appendix.

Because I could not stop for Death,
He kindly stopped for me;

Emily Dickinson

Chapter 7

Rectangularization process of the survival curve

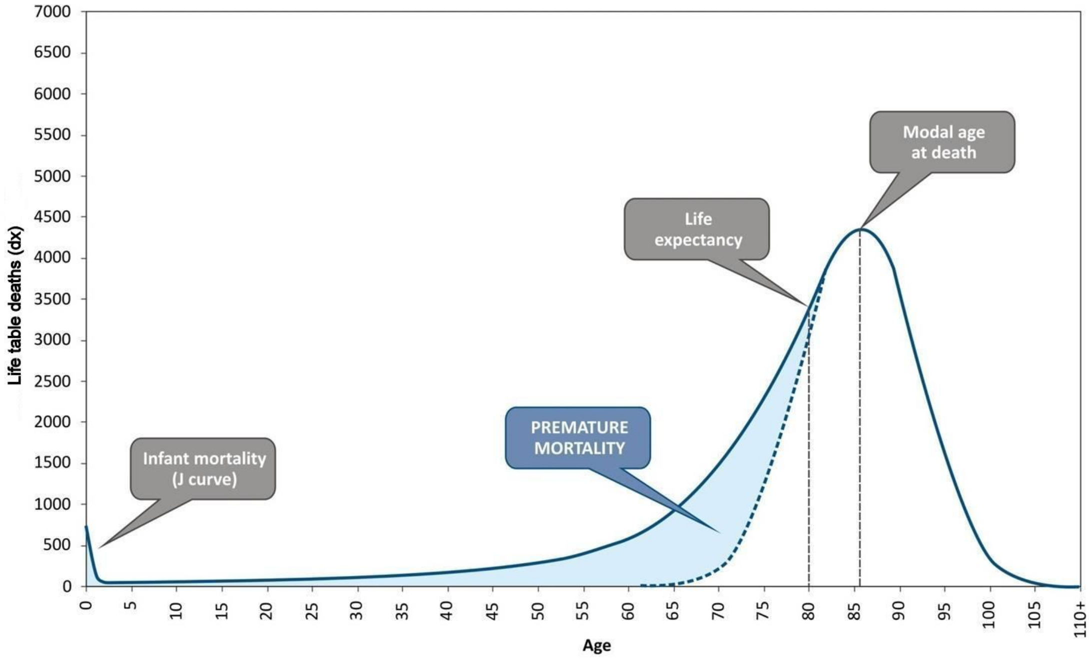
As was mentioned in the previous parts of the Thesis, the study of the rectangularization process could be taken as one possible way to finding the answer whether some limit of the human life span exists. Generally speaking the rectangularization of the survival curve, or compression of mortality, would signify such a theoretical limit. In accordance to that this concept was also integrated to this Thesis. So as the process could be studied or evaluated, some indicators or measures are needed. Those will be described in the text below and used for the basic analysis applied not only to data from the Czech Republic. For the basic analysis data from more European countries were selected – several Eastern and Central European countries (all of them post-communist countries) are compared with a representative of the Western or Northern Europe, with Sweden. For the purpose of this Thesis the issue of the rectangularization of the survival curve was studied using data available from the online Human Mortality Database. The first results of the analysis have been already published at two international conferences (Burcin *et al.*, 2009; Šídlo, Tesárková, 2009).

7.1 Theoretical aspects of the rectangularization process

The shape of the survival curve arose more interest among biologists at first, already in the 1920s. The concept of rectangularization of the survival curve was introduced in the mid-1950s, by Alex Comfort in his book *The Biology of Senescence* published in 1956 (Cheung *et al.*, 2005; Rose, 2007). The theme was then discussed by many other authors, but became known mainly thanks to James F. Fries and his article *Aging, Natural Death and the Compression of Morbidity* from 1980 (Fries, 1980). Fries assumed that the process of rectangularization when the survival curve becomes more and more rectangular is a consequence of a decline in mortality, when more people survive to higher ages. Their deaths are then concentrated in later life because Fries also assumed a relatively fixed genetically determined upper limit of the life expectancy around 85 years of age (Fries, 1980, Cheung *et al.*, 2005).

The base of the theory of the rectangularization process is the study of Lexis from the late 19th century, focusing on so-called normal length of life using Quetelet's model of "average person" (Cheung *et al.*, 2005). Based on that, the modal age at death became the basis of many considerations tied to the theory of rectangularization. According to Lexis, deaths associated with aging and age are centered and normally distributed around the modal age at death. Lexis connected those deaths, which do not correspond to a normal distribution around the modal age, to either early mortality of children or the premature deaths, which occur in relatively young age between infant mortality and deaths that can be associated with high age and senescent (Cheung *et al.*, 2005; Canudas-Romo, 2008; Kannisto, 2000; Burcin *et al.*, 2009; Šídlo, Tesárková, 2009; see Figure 43).

Figure 43: Illustration of the normal distribution of deaths around the modal age at death according to Lexis and premature deaths (real data used as the base, Sweden, females, 1985–1989)

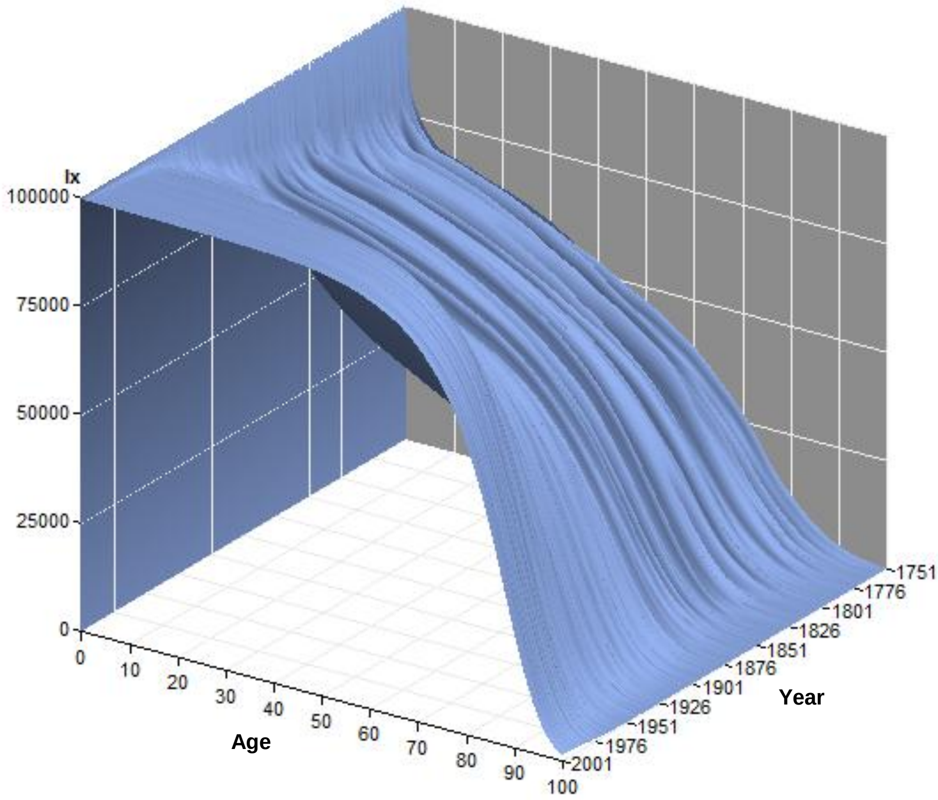


Source: Šídlo, Tesárková, 2009

The process of rectangularization can be viewed directly through a survival curve, but also indirectly through the curve of distribution of deaths in the life table. It was shown that the rectangularization of the survival curve is accompanied by decreasing variability of ages at death around the modal age at death (the process of mortality compression) (Fries, 1980; Cheung *et al.*, 2005; Kannisto, 2000). Over time, however, in demographically developed countries there occurred not only a concentration of deaths around the modal age, but also an increase of the attainable age, and thus a kind of shift of the survival curve to the right. This shift does not reflect more intensive rectangularization or increasing concentration of deaths around the modal age, that is why some authors speak rather about a “derectangularization” (e.g. Yashin *et al.*, 2001).

Basic analysis of the rectangularization process of the survival curve or compression of mortality can be made using the appropriate graphic and illustrating methods. It seems practical in this case to use the three-dimensional graphs. The graphs, which allow monitoring the development of the phenomenon in time and also the distribution of intensity by age, can well illustrate the historical development.

Figure 44: Development of the survival curve, Sweden, females, 1751–2001

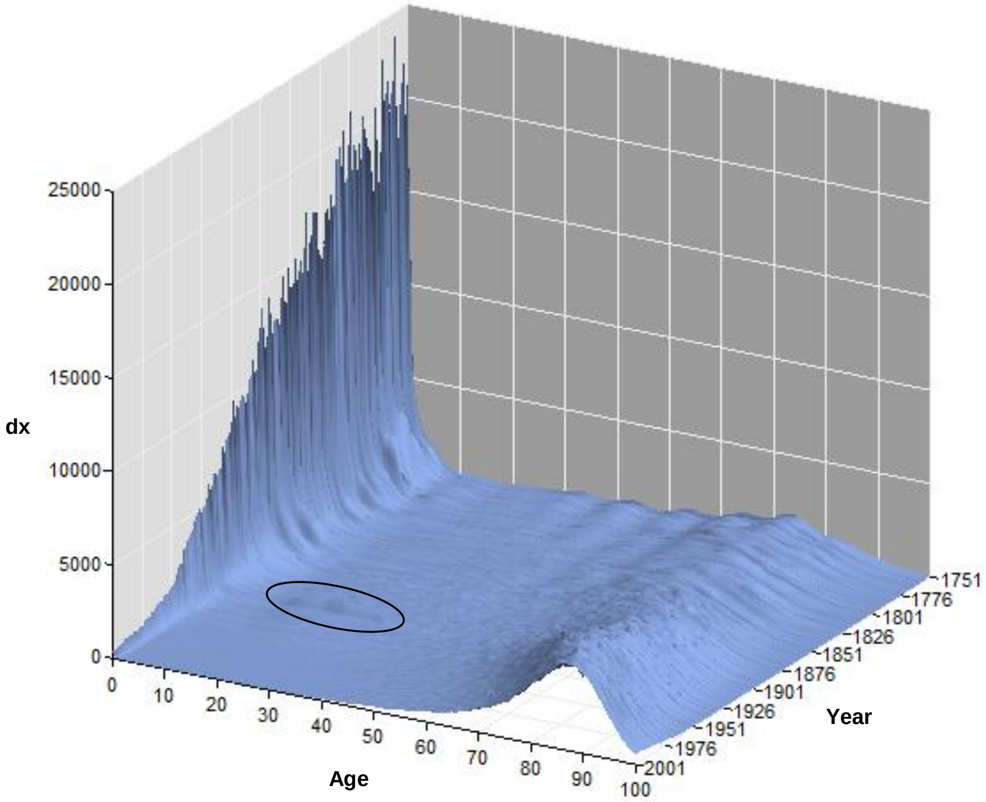


Note: Prepared in SAS 9.2 software
Source of data: Human Mortality Database (2010)

In the Figures 44 and 45 the development of the survival curve and distribution of table deaths for population of Swedish females is shown. This population was chosen on purpose because of the necessary data availability in a sufficiently long time series. Thanks to the three-dimensional graphs it is possible to see the improvement in the mortality rates in the past, as well as significant impact of irregular fluctuations and mortality crises mainly before the World War II. The last visible fluctuation that affected the development of the two curves occurred in 1918 and was due to the so-called Spanish flu epidemic. Particularly in the second half of the 20th century there could be seen a smooth development without sharp changes, and approximation of the shape of the survival curve to almost a rectangle, when a large share of the generation survive to a relatively high age and then die within a relatively narrow age interval.

Even in the figure of the density of table deaths (Figure 45) it could be seen significant progress during the given period. Primarily, at first sight there is apparent a sharp drop of numbers of deaths at the very beginning of life, i.e. during the first years after birth. The last major fluctuation occurred in 1918, again due to the Spanish flu. As marked in the Figure 45 by the ellipse, the epidemic significantly affected people at younger ages, ca up to 40 years. But what is more important for changes of mortality at the highest ages, there is a continuing concentration of deaths around the modal age, which itself is also increasing during past decades. This maximum of the curve of the distribution of table deaths not only shifts to higher ages, but also the deaths are less variable around the modal age. This is the evidence of the assumptions that are associated with the rectangularization theory (Fries, 1980).

Figure 45: Distribution of table deaths, Sweden, females, 1751–2001



Note: Prepared in SAS 9.2 software
Source of data: Human Mortality Database (2010)

Graphical comparison of changes in the shape of the survival curve or curve of the distribution of deaths is just one of the possible methods how to illustrate and analyze the process of rectangularization. It is rather the simplest of them. For more detailed study of this process and for a possibility of comparisons in space and time many specific indicators have been developed. These indicators due to their different design often follow various aspects of the same process and, therefore, they are often strongly correlated (Wilmoth, Horiuchi, 1999).

7.2 Indicators usable for the analysis of the rectangularization process

For the analysis of the rectangularization processes there are many indicators developed and they can be divided into several groups according to various criteria. Cheung *et al.* (2005, p. 245) introduced one of the possible classifications, where the indicators are divided into seven groups depending on what the particular indicator measures or on what aspect of the changes of the survival curve it is mainly focused:

1. They defined a group of “central longevity indicators” characterized by a modal age, median age, or life expectancy.
2. They did not include any indicator to a group called “horizontalization indicators”.
3. On the contrary, most of the commonly used indicators fell to the group of “indicators of concentration and verticalization” (e.g. various ways of calculations of the standard

deviations of life spans or ages at death, the popular indicators of the so-called C-family, Keyfitz's index of entropy).

4. The fourth group consisted of the so-called "rectangularization indicators", which include the indicators of, for example, the fixed or moving rectangle.
5. They included indicators characterizing the maximum life span and longevity to the fifth group called "maximum longevity indicators".
6. In the other group of indicators are those describing the whole survival curve (above all i.e. percentiles), these indicators are called the "mapping indicators".
7. And finally in the seventh group ("other indicators") there were the rest of the indicators such as the coefficient of variation or the Gini coefficient which were not included to any before mentioned group.

It is also possible to use another, more methodological, way of distinguishing the indicators. Based on this point of view we can define three basic groups of indicators. This categorization will be used in this chapter:

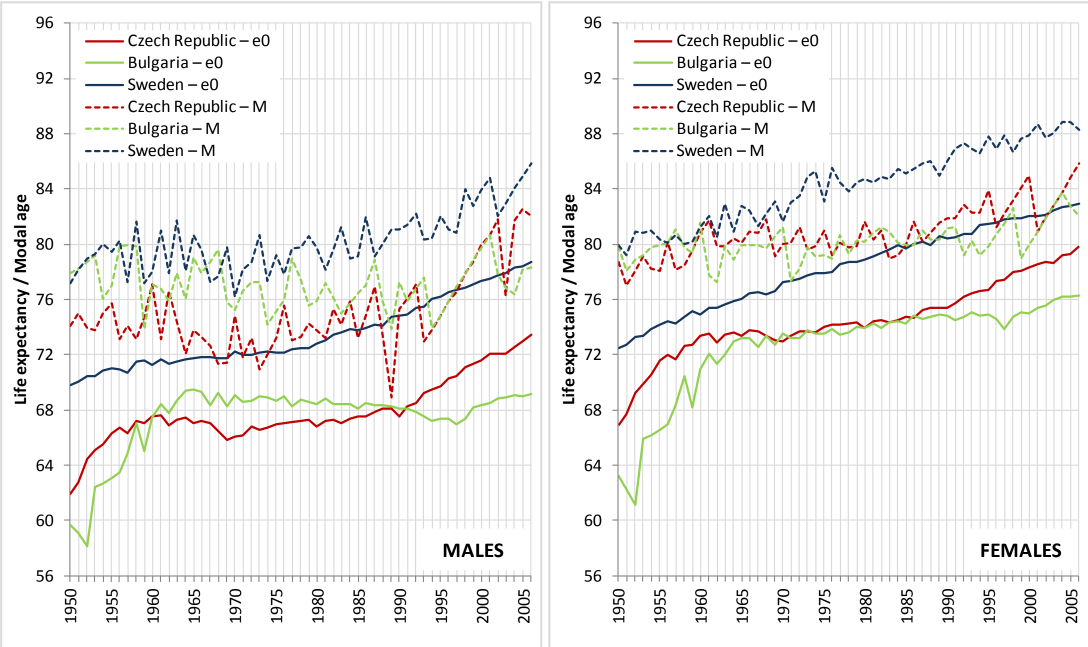
1. The first group of indicators as defined for the purpose of this Thesis is filled by the pure demographic indicators which are commonly used within the demographic analysis and demographers usually take these indicators as to be the most important ones which can easily characterize the whole survival curve (e.g. life expectancy, modal age, etc.).
2. The second group consists of several indicators based on more statistical or mathematical calculations and they are derived from the basic demographic indicators. It is possible to mention, for example, indicators of variability and intensity that are computationally linked to the indicators included in the first group (e.g. standard deviation of age at death, etc.).
3. The last group could then be formed by those indicators that are directly designed to measure the intensity of rectangularization, or by the indicators, although specially designed for other purposes (such as measurement of inequality in income distribution in economics in the case of the Gini coefficient), which are also usable for the purpose of calculations related to the rectangularization process.

7.2.1 Basic demographic indicators

In the first group of indicators, i.e. the group of the basic and in demography commonly used ones, there not only the mode can be included but also for example the median, life expectancy and frequently used percentiles and quartiles, and thus also an indicator of interquartile range.

Indicators such as the modal age or life expectancy are traditionally used characteristics of the mortality process. The life expectancy is standard and most frequently reported outcome of the life tables. In fact, it is an expected value of a random variable which could be called the length of future life of a person at an exact age x; in interpretation it is a number of years that the average person at the exact age of x years is likely to live more. Most frequently the life expectancy of a person at the exact age of 0 years (life expectancy at birth) is used. The modal age at death represents basically the so-called normal life span, and this is the age at which the greatest number of people died from the analyzed population (Koschin, 2002).

Figure 46: Development of the life expectancy at birth and of the modal age at death, selected countries, by sex, 1950–2006



Note: e0 = life expectancy at birth; M = modal age at death

Source of data: author's calculation based on Human Mortality Database (2010)

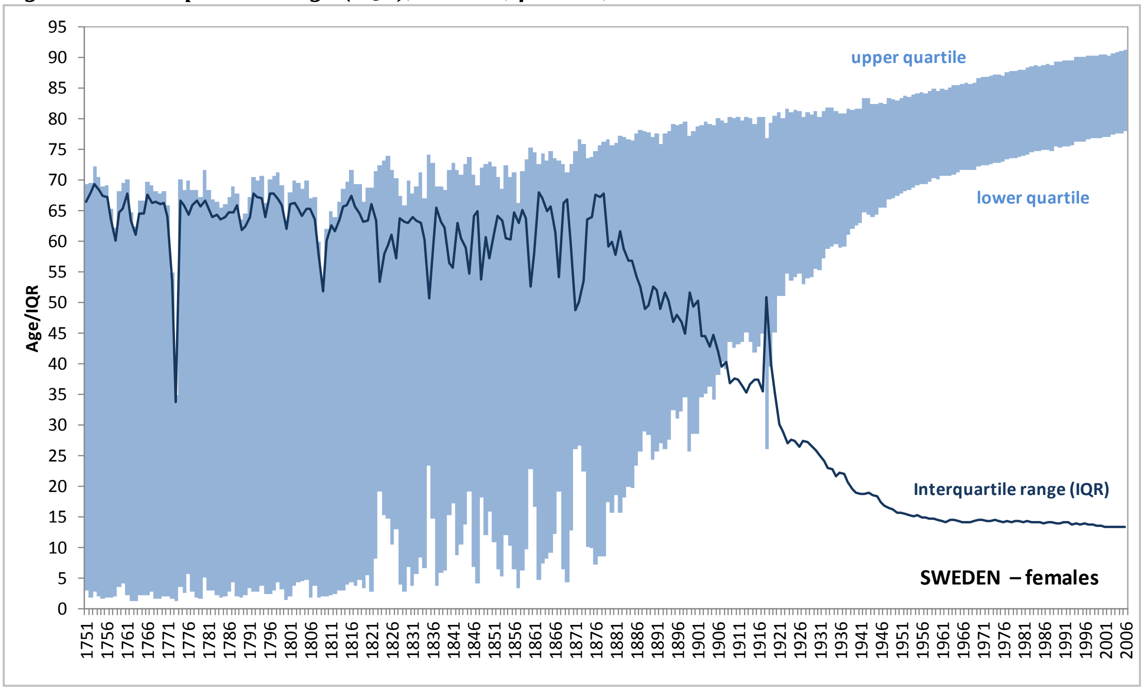
As a disadvantage of the life expectancy it can be considered the fact that in terms of its design it reflects mortality of all ages. It is therefore quite significantly affected by changes in mortality early in life which is already very low in demographically developed countries. In contrast, the modal age at death reflects changes in mortality at higher ages. Because of that, this indicator is still more and more often used in demographic analysis. Its disadvantage is relatively high sensitivity to random variations so its development in time is not smooth. In the Figure 46, there the development of both those indicators in three selected European countries since World War II is illustrated – the situation in the Czech Republic is compared with Sweden (as a representative of the Western Europe) and Bulgaria (representing the post-communist countries in the Eastern Europe). It is evident that in all those three countries the improvement of mortality rates during the period occurred. However, significantly better is the position of Sweden and different development of the Czech Republic and Bulgaria after 1990 cannot be overlooked too. While in the Czech Republic rapid improvement without any stagnation was started at the end of 1980s, in Bulgaria some worsening could be seen (in case of males) or at least stagnation (for females) after the change of the political regime. From this basic illustration it is clear that the development in post-communist countries cannot be taken as a universal one. That is the reason, why in later analyses usually more Eastern European countries were taken into the account.

Another frequently used indicator of the demographic analysis is the interquartile range (marked as IQR). It is the difference of lower and upper quartile of ages at death in a given population or generation. Interquartile range is the indicator offering a lot of other ways of usage and of deeper analysis than simply an assessment of its development in time, as will be shown later.

For contemporary populations, there is no reason to calculate the interquartile range from values of numbers of survivors above a certain age because the infant mortality is already not playing any important role in the overall mortality. However, in historical or less-developed populations the lowest ages are commonly omitted in its calculation (Wilmoth, Horiuchi, 1999). Frequently the indicator was calculated for ages at least 10 years. Even in that case when the lowest ages are omitted the same method of calculation is used – the calculation of the appropriate age of the lower and upper quartile and a simple difference of them. However, only the numbers of deaths are considered where the age is above some minimal value (for which the indicator is calculated).

If we decide to calculate the interquartile range above a certain age (as described in previous paragraph), it is logically possible to calculate it for each particular age – i.e. it could be calculated for the entire age range (i.e. above age 0), or for age 1 (i.e. above age 1), for age 2 (above age 2), etc. In such case a lower index has to be added to the label of the indicator identifying the lowest age entering the calculation, IQR_x. Therefore, the interquartile range could be studied not only from the long term perspective (i.e. evaluated its development in time), but also its development according to age can be analyzed and then possibly also the change of the dependence on age in time could be studied.

Figure 47: Interquartile range (IQR), Sweden, females, 1751–2006



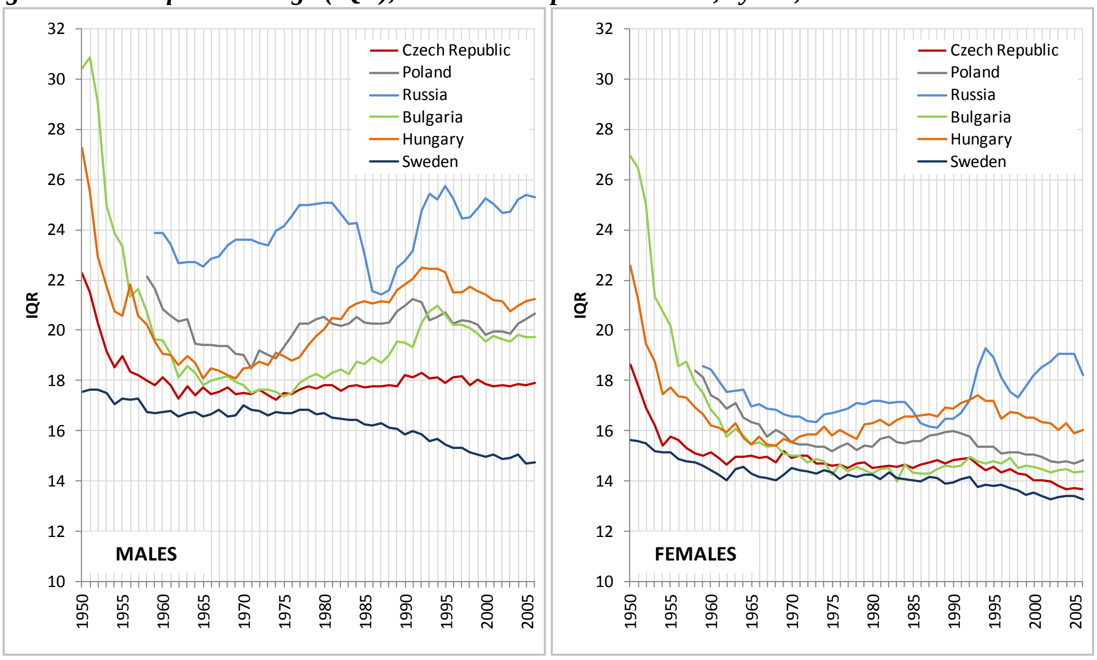
Source: Šídlo, Tesárková, 2009

As in the development of many other indicators, also in the case of the interquartile range many considerable fluctuations of its values could be seen in the past. This is particularly well observable in the case of Sweden (Figure 47), for which we have the longest time series of data. Continuous decline in values of this indicator is evident from about the last quarter of the

19th century, since when the last significant spike in the further development was caused by the epidemics of Spanish flu in 1918.

Also there is not any problem with the interpretation of this indicator. This is essentially the age interval in which 50 % of people born in one generation will die if we do not consider 25 % of deaths in the lowest or highest ages (lower and upper quartile). Those are therefore 50 % of deaths which occur at the ages around the median age at death in a given population. If we calculate the interquartile range for different ages (IQR_x), the result is again an age interval within which 50 % of people die but not 50 % from the whole generation, but from the people alive at the exact age *x*. According to the graph (Figure 47), in Sweden until the last quarter of the 19th century 50 % of people deceased in the period lasting almost 65 years, one quarter deceased before the age of about 5 years and one quarter at the age over 70 (i.e. 65 + 5) years. Currently, one quarter of Swedish females die in the lower quartile which lasts up to the age around 75 years, a quarter dying in the upper quartile then survives to the age above 90 years, the interquartile range (where 50 % of people in the considered generations die) is then only approximately 15 (i.e. 90–75) years of age. Due to significant differences among Eastern European countries there are more of them involved into the analysis (as was mentioned above): the Czech Republic, Poland, and Hungary represent the post-communist countries in the Central Europe with its specific mortality conditions and life-style of the population, Russia and Bulgaria represent the Eastern part of Europe.

Figure 48: Interquartile range (IQR), selected European countries, by sex, 1950–2006



Source of data: author’s calculation based on Human Mortality Database (2010)

If we focus the attention only on the period after World War II, practically no major fluctuations of this indicator for Swedish females could be seen. Its values gradually decreased, but not very rapidly, for males the rate of decline has increased from about half of the 1970s. In the case of the Czech Republic, there a sharp decline in the postwar years is obvious and further development for females was roughly similar to Sweden, but the values were higher for ca one year. For males in the Czech Republic any decrease still has not been almost noticeable. But it could mean also that both, the lower as well as the upper quartile, are increasing (see Figure 48).

However, especially for males, there is a significant difference between the Czech Republic and other post-communist countries included in the analysis. The situation in Poland, Bulgaria, Hungary and Russia could be characterized by the increase of the indicator for males ca from the 1970s (in Russia already from the 1960s) while in the Czech Republic the same time period could be characterized rather as a stagnation. For females the situation was more stable in all the involved countries.

After the change of the political regime in the Eastern and Central European countries the situation of mortality got even worse in some of them – e.g. Bulgaria and Russia. In Hungary and Poland the indicator started to decrease with a slight sign of worsening only in the latest years. For females the decreasing tendency after the end of 1990s could be seen in all the involved post-communist countries except for Russia.

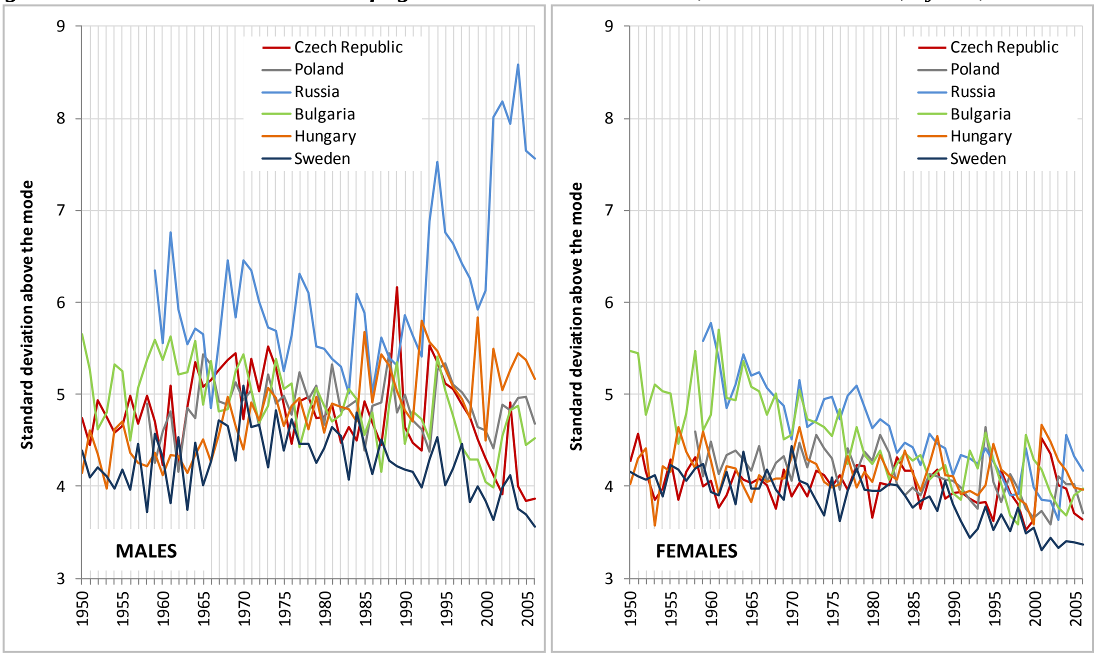
The development in Russia was rather unique during the studied period. Mortality conditions in this country got worse already from the 1960s (visibly for males) what is usually connected with the worsening of economic situation in the country but also as a presage of the incoming crisis of the Soviet system. Sometimes this mortality development during the 1960s and later is taken as a warning "that proceeded that of any of the conventional macroeconomic indicators" (Shkolnikov *et al.*, 2004, p. 31). The fluctuation during the second half of the 1980s could be related to the anti-alcohol campaign during years 1985–1987 and return to alcohol consumption at the beginning of the 1990s. Moreover, this time was affected also by the economic collapse and the start of the transformation of the political system (Shkolnikov *et al.*, 2004). It seems from the Figure 48 that the economic crisis at the end of 1990s affected more females than males. In general, Russian demographic trends at the end of the 20th century would be sufficient theme for an individual analysis (Gavrilova *et al.*, 2002; Kocová, 2009). The specific demographic development (or rather mortality development) in Russia will be obvious also in all the following graphs and analyses within this Thesis.

7.2.2 Indicators derived from the basic ones and its application

The information provided by indicators from the first group is easily expandable with a few simple calculations. Thus it is possible to calculate the variance of ages at death or coefficient of variation. If the calculations follow the mode of ages at death it can be specified the average age at death above the mode or the variance of ages at death above the mode. If the calculated quartiles are used, it is possible to estimate the average age at death or the variance of ages at death, for example, only in the highest quartile.

For example, the standard deviation of ages at death above the mode changes much more slowly (except of several extreme changes during the 18th century) than most of the other indicators (Figure 49). The indicator is slightly decreasing over time – the standard deviation above the modal age thus slightly decreases, which spoke in favor of the process of rectangularization, or compression of mortality. However, it is important to mention that even the modal age increases, so actually we can study the standard deviation of ages at death which occur in a still shrinking age interval. At the beginning of the 21st century we can already see the values stabilizing or even slightly increasing. For Russian males it is the result of unfavorable demographic situation as was mentioned above. On the other hand in case of Japanese females, which were not included to the presented charts, or of Swedish females that would indicate the shift of the end of the survival curve to higher ages (to the right). It will be described later.

Figure 49: The standard deviation of ages at death above the mode, selected countries, by sex, 1950–2006

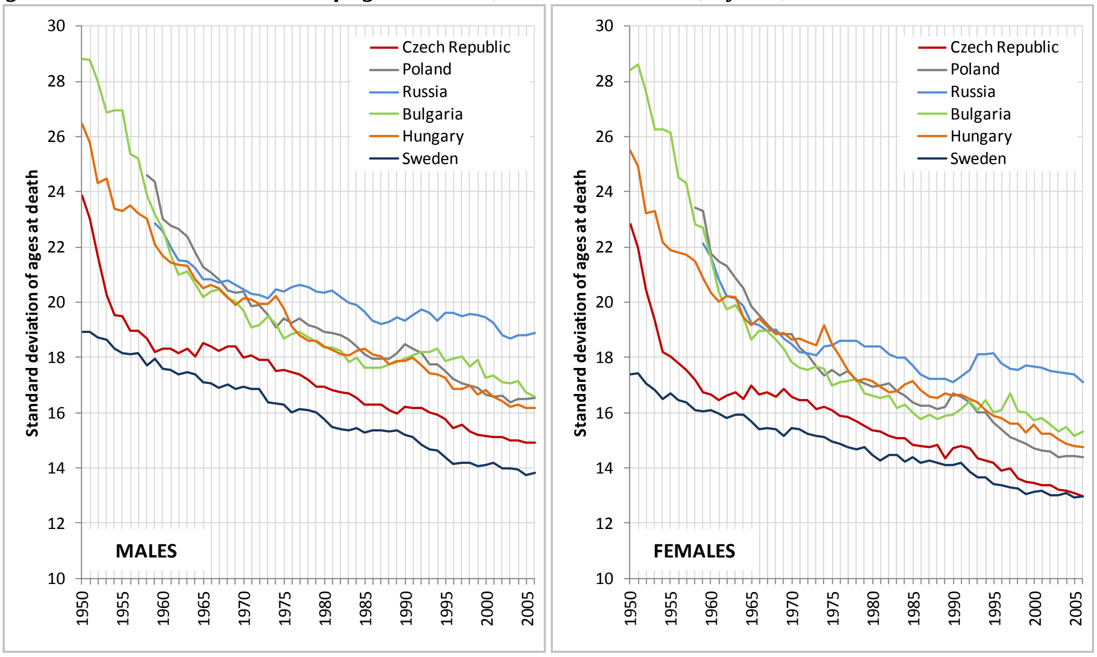


Source of data: author's calculation based on Human Mortality Database (2010)

In contrast, the standard deviation of ages at death calculated for the entire age range in time decreases much more significantly as a result of the impact of changes in mortality at lower ages on the value of this indicator. At the same time, greater differences between countries are more visible. While in Sweden was a gradual decline from the end of the 19th century, countries like the Czech Republic and other post-communist countries experienced a sharp decline after the World War II. Currently, the rate of decline decreases and in some states the stagnation of values of the indicator is also apparent.

From the Figure 50 it could be seen that this indicator was not affected by the Russian mortality crisis as could be expected. On the contrary the indicator of standard deviation above the mode (Figure 49) was affected in case of Russian males. It could mean that the variability of ages at death at higher ages was influenced by the crisis but the overall variability rather stagnated over time. The reason could be found by the analysis of age-pattern of mortality when the increase of the overall variability caused by the worsening of the situation of adults was possibly at least partly compensated by the improvements of the infant and child mortality (this assumption is based on the analysis made by Kocová, 2009).

Figure 50: Standard deviation of ages at death, selected countries, by sex, 1950–2006



Source of data: author's calculation based on Human Mortality Database (2010)

7.2.3 Specifically designed indicators

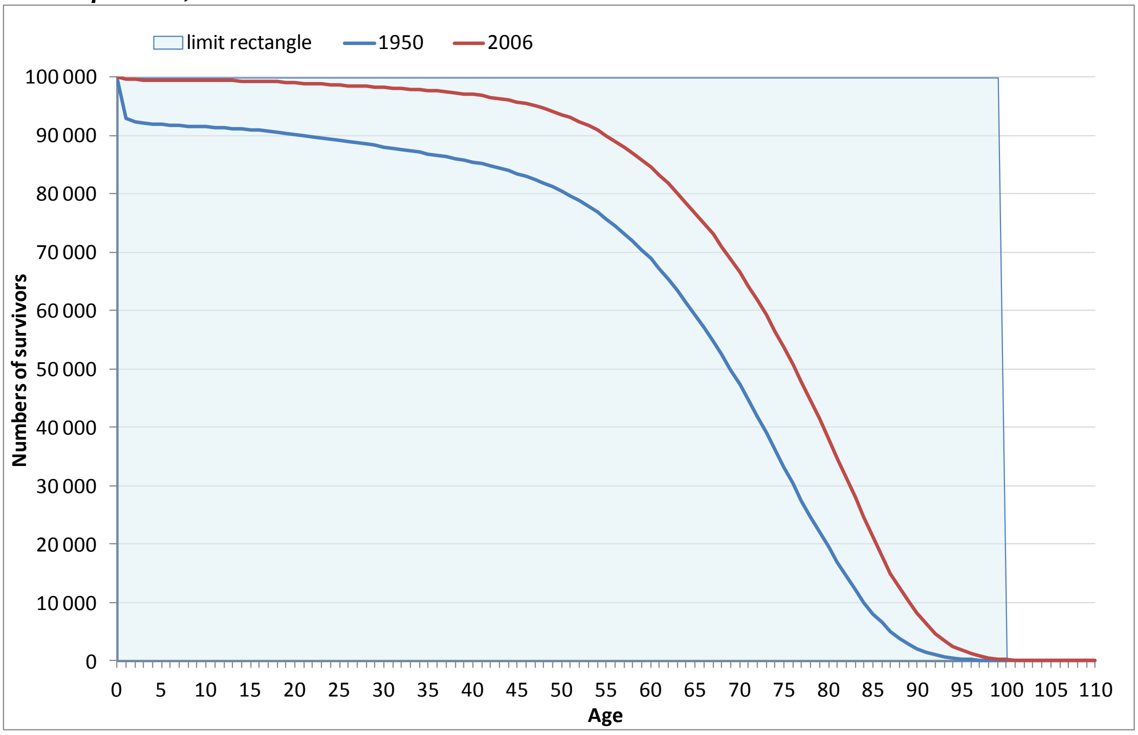
The rest of the indicators used for the analysis can be classified into the third group, indicators that have been specifically designed for illustration of the rectangularization process or its particular aspects. In this group, there would be included the indicators of longevity, called fixed or moving rectangle, the Gini coefficient, index of entropy (or Keyfitz's H), or indicators from the so-called C-family (Wilmoth, Horiuchi, 1999; Kannisto, 2000; Cheung *et al.*, 2005).

Fixed rectangle

Calculation of fixed and moving rectangle (Wilmoth, Horiuchi, 1999) is based on a simple comparison of the area under the survival curve and the area of theoretically possible (limit) rectangle (Figure 51). The area of this limit rectangle is determined by the length of its one side which is equal to the number of survivors at age 0 and the other side, which in case of the fixed rectangle is given (as e.g. age 85 or 100). For the moving rectangle the length of the second side corresponds to such an age, to which only a certain proportion of the original cohort survives (for example 1 / 1000 or 1 / 10000 of the original population) (Wilmoth, Horiuchi, 1999).

The area under the survival curve can be calculated as the value of its integral. In case of a discrete form of expression related to the number of survivors, as common in the life tables, a solution can be obtained via the so-called rectangular method. In this method the area under the curve is expressed as the sum of areas of particular rectangles, into which the area was divided. Heights of these rectangles are the numbers of living in the middle of the age interval included in the particular rectangle and the width is then equal to this age interval (Figure 52). Thus, in case of one-year age intervals the calculation is simplified to the sum of individuals living at the completed age *x*. Of course, with the ongoing process of rectangularization, if we understand it as an approximation of the shape of the survival curve to the rectangle, the values of fixed and moving rectangles are approaching one.

Figure 51: Illustration of the fixed rectangle indicator at age 100, Czech Republic, males, data of the survival function, 1950 and 2006

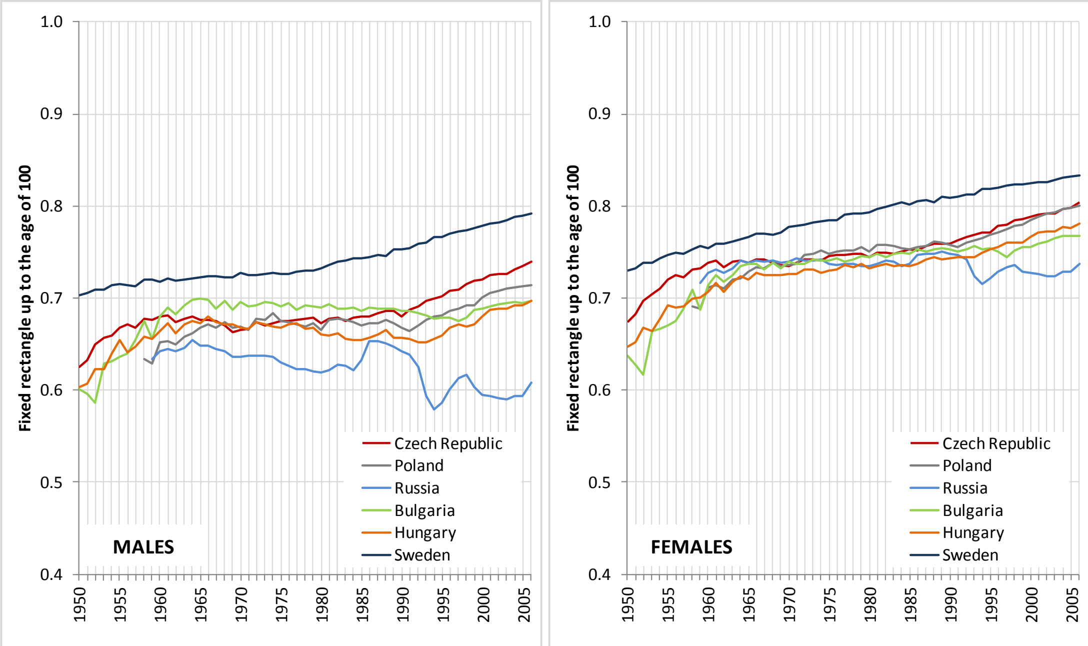


Source of data: author's calculation based on Human Mortality Database (2010)

Figure 52: The rectangular method for calculation of the area under the survival curve

Source of data: author's calculation based on Human Mortality Database (2010)

Figure 53: Indicator of fixed rectangle with the limit value at age 100, selected countries, by sex, 1950–2006



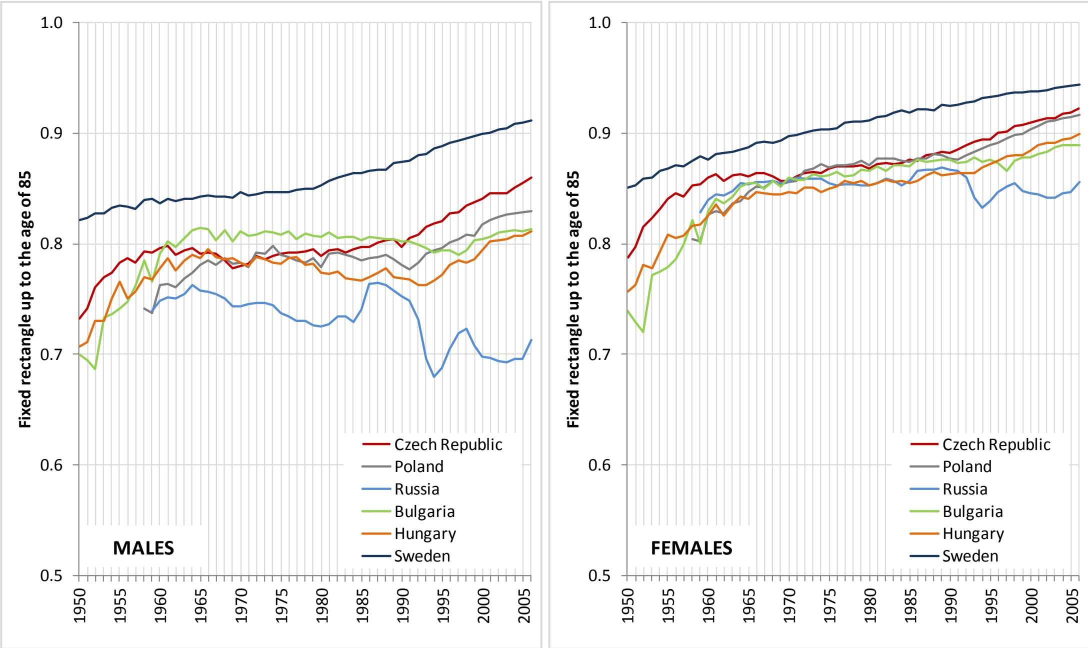
Source of data: author's calculation based on Human Mortality Database (2010)

For the fixed rectangle with the limit value at the age of 100, there it is noticeable a different development in individual countries selected for comparison (Figure 53). From the selected post-communist countries the Czech Republic keeps the best position, where the rapid convergence to the Swedish values is apparent, especially after 1990. For females the same could be said for Poland. On the other hand, for Russian females any long-term improvement could be seen already since the 1960s, the situation of Russian males is getting even worse in some periods. It is a consequence of the factors mentioned above – life style of the population, the economic and social development of the country.

A situation where the value of the indicator of fixed rectangle become equal to one is practically impossible, without any doubt, there always remains some variability of ages at death there. Nevertheless its values are relatively close to one, reaching over 0.8 (in case of Sweden), suggesting a relatively high level of rectangularization and relatively small area where this process can proceed in the future. How far this indicator can rise remains a question. For example, for Swedish females (or for countries such as Japan) it is already visible some slowing of the growth and it is likely that the value of the fixed rectangle theoretically may increase approximately to the value of around 0.9.

While for the fixed rectangle with the limit value at age 100 it was necessary to assume the presence of some variability in the ages at death, for the fixed rectangle limited to the age of 85 years this variability can be expected to be lower. Because of the improvements in mortality in developed countries it is theoretically possible that the variability of the ages at death within 85 years of age will indeed decline to some minimal values. Values achieved for this indicator are really much closer to one, again primarily for Swedish females. Again, almost no long-term improvement of the indicator is evident for Russian females and males (Figure 54).

Figure 54: Indicator of fixed rectangle with the limit value at age 85, selected countries, by sex, 1950–2006



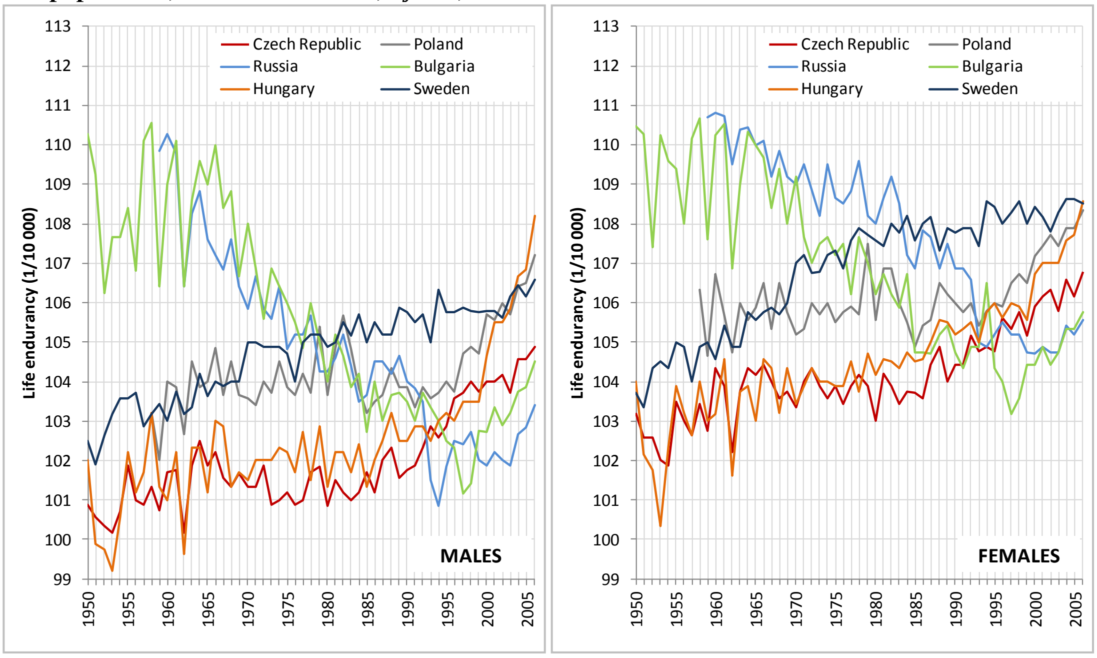
Source of data: author's calculation based on Human Mortality Database (2010)

Indicator of life endurance

Indicators of life endurance ("longevity") are mainly focused on the monitoring of growing probabilities of people (and greater proportions of population) to survive to still higher and higher ages. In this way it can be characterized as an indicator of age, to which theoretically survives the last person from 10000 or 100000 people from the original cohort (Cheung *et al.*, 2005). For demographically developed countries with traditionally low levels of mortality (as Sweden for example), this indicator does not show any significant increase in the long-term perspective. Its values exceeded 100 years already for the data from the 18th century. This fact would support the assumption that the maximum length of human life has its fixed limit and that it cannot grow indefinitely. However, relatively high values for Sweden during the 18th and early 19th century are rather the result of inaccurate data than extremely favorable mortality conditions. Nevertheless, the data does not confirm any tendency to greater increases of the values of this indicator.

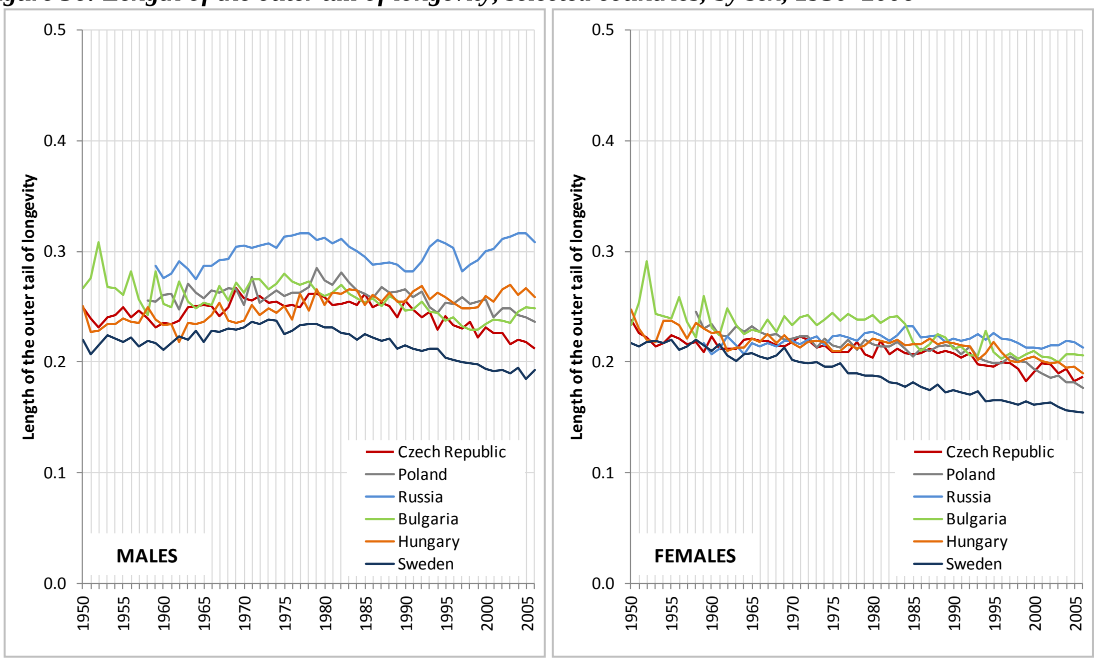
Higher rate of increase was shown only for limited time periods, as is evident for example in the Czech Republic or Hungary, or in Japan (as a representative of non-European countries, not shown in the graphs), where there was a significant and rapid increase after the World War II; in recent years the values are rather stagnating or slightly growing in the Eastern European countries (Figure 55). However, it must be said that this indicator is calculated from data published in the Human Mortality Database, these data are smoothed by the Kannisto method and that is why the results could be influenced (Wilmoth *et al.*, 2007).

Figure 55: Indicator of life endurance, the age to which survives at least 1/10 000 of the original size of the population, selected countries, by sex, 1950–2006



Source of data: author’s calculation based on Human Mortality Database (2010)

Figure 56: Length of the outer tail of longevity, selected countries, by sex, 1950–2006



Source of data: author’s calculation based on Human Mortality Database (2010)

For the selected group of countries subjected to the analysis, there some specific development was manifested mainly for Russia and Bulgaria, for both sexes. While in other countries, this indicator slightly increased (as in Sweden) or the growth temporarily stagnated (the Czech Republic, Hungary, Poland – especially in case of males), the situation in Russia and Bulgaria was opposite and values of the indicator of life endurance showed rather a decline from the 1960s up to 1990s. However, inaccurate data in the post-war period may at least partially explain the revealed fact.

It is unlikely, that situation in Russia and Bulgaria was significantly better than in Sweden in the post-war period.

Length of the outer tail of longevity

Another indicator focused on mortality at the highest ages is the length (or rather, variability) of the final part of the survival curve (the indicator is called "length of the outer tail of longevity") (Cheung *et al.*, 2005). The indicator is defined as the difference of the highest age attained in the population and the age to which the oldest 10 % of people survived divided by this age (*ibid.*). Values of this indicator along with improving mortality are decreasing. It indicates that the variability of ages at death at the highest ages rather decreases and the last surviving 10 % of the original population die within still a shorter period of time. This result will be confirmed later in the analysis.

Gini coefficient

The Gini coefficient is an indicator known primarily from economics. There it found the greatest applicability especially in the evaluation of income inequality in society. It is based on the Lorenz curve, which graphically shows the uneven distribution. Gini coefficient in its basic form is defined as the ratio of the area between the curves (the real one and the limit one) and the area under the limit curve (Wilmoth, Horiuchi, 1999). Its definition is very similar to the definition of a fixed or moving rectangle. Its values vary between zero and one, where zero indicates a situation where the maximally possible attainable curve and the curve of the real distribution merge into one.

As the process of rectangularization or compression of mortality continues values of the Gini coefficient gradually decline, as the actual real survival curve approaches the "ideal", indicating a possible limit pattern with minimal variability of ages at death. Again, it is not possible that the indicator is absolutely equal to zero. In such a situation theoretically everyone would live an equally long life and die at the same age, so the variability of ages at death would be zero. Nevertheless, the values of the coefficient reduce in the long-term perspective, however, in recent years their development clearly tends to stagnation at a level slightly below 0.1. The coefficient rapidly declined mainly after the World War II – for example, in Bulgaria or the Czech Republic.

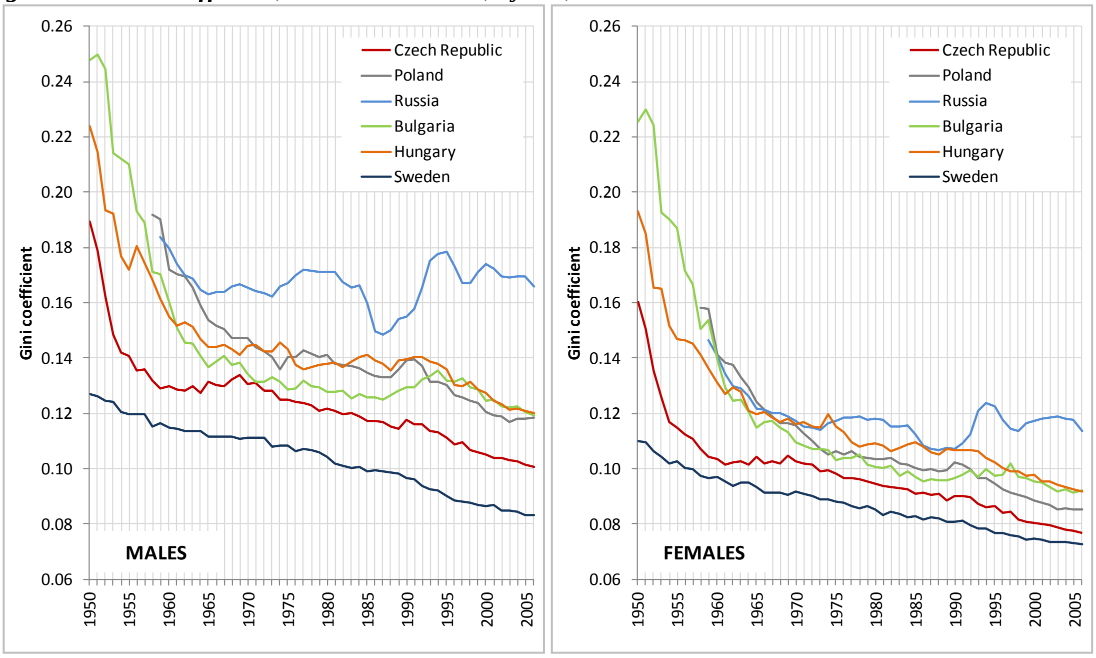
The Gini coefficient was calculated as (Wilmoth, Horiuchi, 1999):

$$G = 1 - \frac{1}{e_0} \sum_{x=0}^{\omega} \left(\frac{l_x}{l_0}\right)^2,$$

where x is age, ω is the lowest age to which no one from the original population survives, e_0 is the life expectancy at birth, and l_x are the values of the survival curve at the exact age x.

Within the analyzed group of countries, there it is still apparent the difference between Sweden and the displayed post-communist countries (Figure 57). Low values of the Gini coefficient (especially for Swedish females) again provoke the question of how the future development might look – how long we can expect the indicator to decrease, how fast this decline may be, etc. Situation in Russia is the opposite one to the example of Sweden, particularly for Russian males. Gini coefficient reflects very well the mortality crises, because values of this indicator react very sensitive to the ongoing changes in society that are naturally reflected also in the level of mortality in the country (Gavrilova *et al.*, 2002; Shkolnikov *et al.*, 2004; Kocová, 2009).

Figure 57: Gini coefficient, selected countries, by sex, 1950–2006



Source of data: author's calculation based on Human Mortality Database (2010)

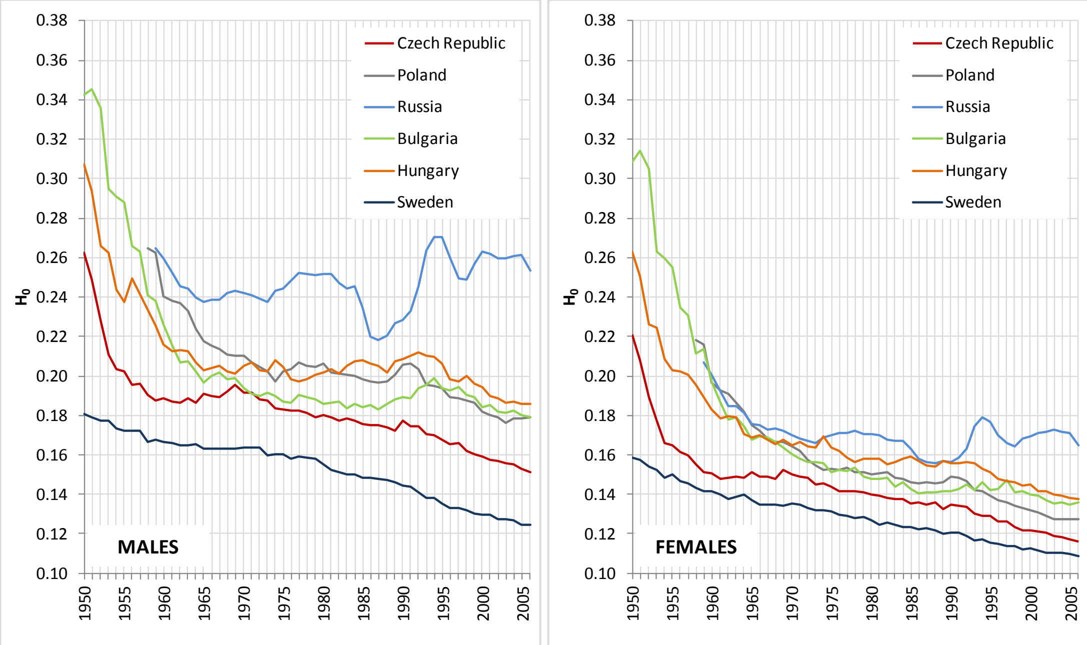
Keyfitz's index of entropy

One of the most frequently used and most important indicators applicable in the analysis of the development and overall level of mortality compression and rectangularization is the index of entropy (or Keyfitz's H). The possibilities of this indicator, however, are much broader. The indicator is described for example in the article from Nagnur (1986). The Keyfitz's index, labeled as H_a, is defined as follows (in some cases, the numerator of the expression is used as an independent indicator):

$$H_a = -\frac{\int_a^{\omega} \log l(x) * l(x) dx}{\int_a^{\omega} l(x) dx},$$

where a is the age for which the indicator is calculated, ω is the highest attainable age supposed in the analysis and $l(x)$ is value of a survival function for an exact age x (from the life tables). In the original concept continuous expression of the table functions is used (Nagnur, 1986). In addition to the fact that with the improvement of mortality the values of this indicator are declining in time, the indicator could be used for a different way of interpretation. Keyfitz designed the index in such a way, that it can model the expression of the percentage change in life expectancy at any exact age a, if there is a 1 % decline in the specific mortality rates from age a to ω. To monitor the process of rectangularization in time usually only the indicator H_0 is used, but the expression of the Keyfitz's index is of course possible for all the ages (*ibid.*).

Figure 58: Keyfitz's index of entropy (H_0) at age 0, selected countries, by sex, 1950–2006



Source of data: author's calculation based on Human Mortality Database (2010)

The values of the Keyfitz's index of entropy for age 0 do not show any different trends in time than other indicators (Figure 58). Also by the values of this index it is possible to demonstrate the effects of the Russian mortality crisis, as well as the gradual convergence of values of the Czech Republic to Western and Northern European countries represented here by Sweden. The advantage of the Czech Republic among other post-communist countries involved into the analysis was started already at the end of 1960s (for males) or just after the World War II (for females). If we use the aforementioned possibility of interpreting of this indicator, we can say that the life expectancy for a person at the exact age 0 can theoretically increase by approximately 0.12 % in case of the Czech females if there was a decrease of all age specific mortality rates by 1 %.

C-family

Indicators of so-called "C-family" are also very often used for quantification of the process of rectangularization (Kannisto, 2000). These indicators express the shortest period of age in which the given proportion of the population dies, so e.g. indicator "C50" stands for the shortest period of age in which 50 % people from the original population die. Values of these indicators fall when the variability of the ages at death are decreasing, that is in accordance to the rectangularization process. In the calculation again the lowest ages were often omitted, which was reasonable when a large proportion of people died within the first year (or few years) after birth.

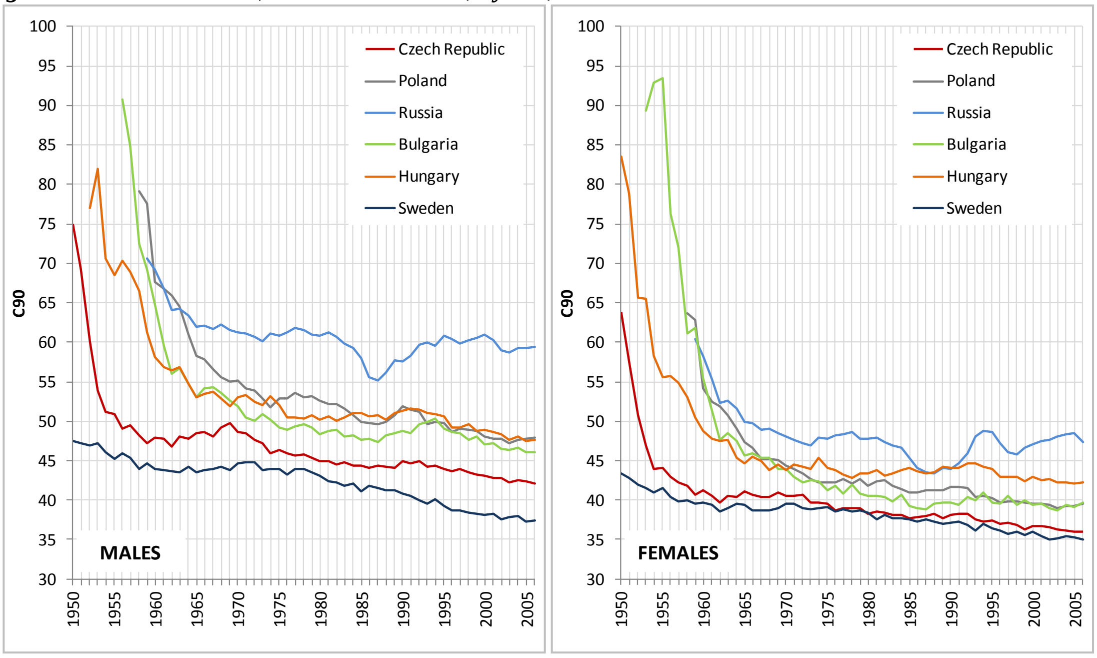
At a time when a large proportion of people live up to very high ages and deaths are relatively concentrated in these highest ages, then the indicator C90 is used very often, i.e. the shortest possible age interval in which 90 % of people from the considered population or cohort die (Kannisto, 2000).

From the shown graphs (Figure 59) it is evident, based on the indicator C90, that the variability of ages at death in the selected countries decreased in time. Only for males (again, with a specific development in Russia) the post-communist countries experienced stagnation, especially in the 1960s and 1970s. The rapid decline immediately after the World War II is not only due to the improvement in mortality in selected countries, as well as due to bettering of the data quality.

Values of the indicator achieved for Czech females are very similar to those of Swedish females, in both these countries ca 90 % of all deaths in the studied populations occur within the period of only approximately 35 years, for men this period is a bit longer. However, in case of females almost the same value for the Czech Republic and Sweden does not signify the same mortality situation. The age interval where 90 % of deaths occur may be the same, but this indicator does not express at which ages this interval lies. A deeper analysis of so-called "normal deaths" can help to reveal not only the length of the interval where most of the deaths occur (in case of normal deaths, where the deaths related to senescence occur) but also the age-localization of those deaths (see below).

From the perspective of the C90 indicator also the situation of males in Russia seems to be relatively stable from ca 1960s except the time of anti-alcohol campaign when probably many premature deaths were avoided and postponed to higher ages so as the interval of 90 % of deaths narrowed.

Figure 59: Indicator C90, selected countries, by sex, 1950–2006



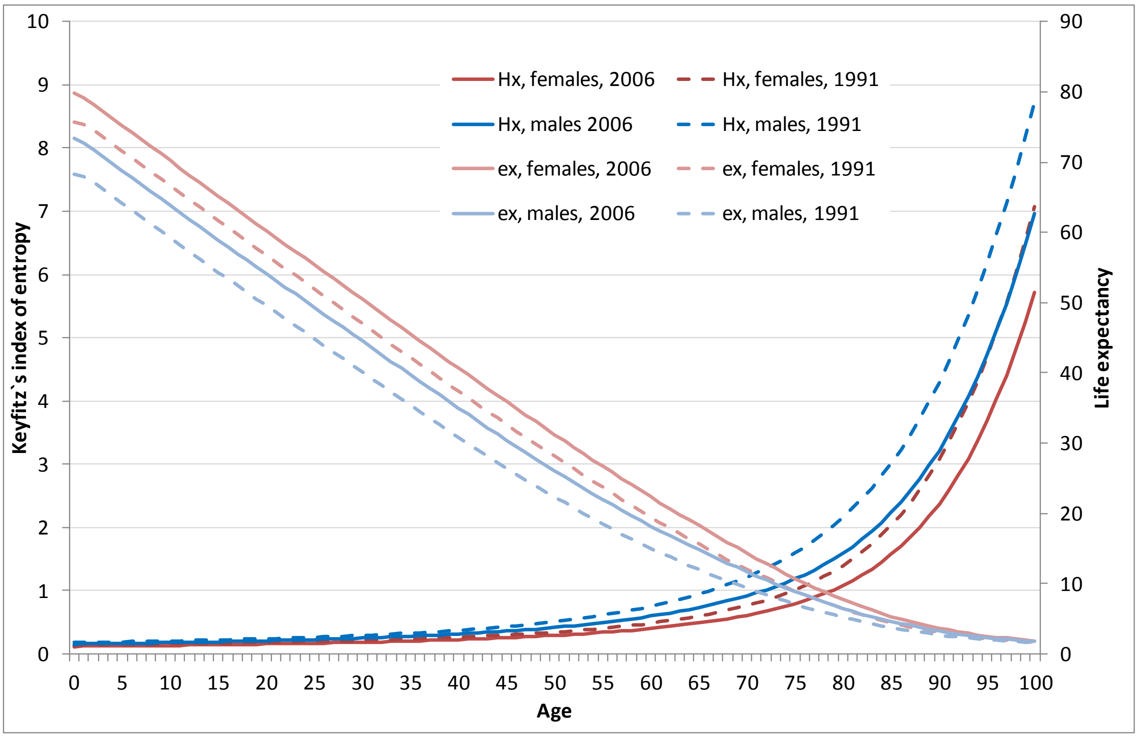
Source of data: author's calculation based on Human Mortality Database (2010)

7.3 Advanced analysis of selected indicators

7.3.1 Wider usage of the Keyfitz's index of entropy

The Keyfitz's index of entropy was introduced in the previous part of the chapter. There the index calculated for the age 0 was used. But this index enables even more detailed analysis. For example much more interesting is the analysis of Keyfitz's index of entropy in dependence on age, in that case the index is calculated not only for the age 0 but for all ages. The Keyfitz's index of entropy has in demographically developed countries a typical age-pattern with low values up to relatively high ages. This means that the potential for possible improvements of the life expectancy at lower ages is already relatively small. In later life the values of the index are increasing, thus also the possibility of improvement of the life expectancy due to decrease of specific mortality rates at these higher ages is increasing. If again we use this way of interpretation of the index, then the decline of age specific mortality rates by 1 % can in the Czech Republic cause an extension of the life expectancy less than 1 % for ages up to about 70 years (as was evident for the value of the index at age 0), however, the same decline in mortality rates may lead to prolongation of the life expectancy by 1 % or more at higher ages (more than 70 years, see Figure 60). The graph also shows that with increases of life expectancy (in the Figure 60 years 1991 and 2006 are shown) the values of the Keyfitz's index of entropy fall at higher ages, thus the potential of further possible increase in the life expectancy decreases especially at those ages.

Figure 60: Keyfitz's index of entropy and life expectancy according to age, Czech Republic, by sex, 1991 and 2006



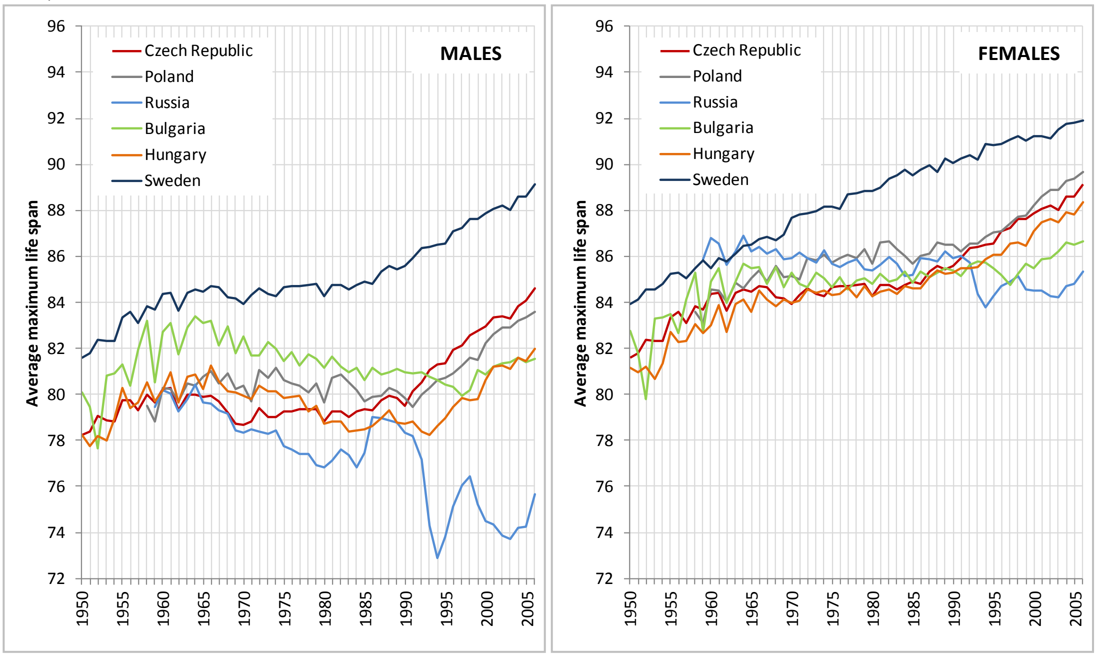
Note: Hx = Keyfitz's index of entropy at age *x*, ex = life expectancy at age *x* (right axis)
Source of data: author's calculation based on Human Mortality Database (2010)

On the basis of the introduced alternative way of interpretation of the index (as a possible increase in the life expectancy), it is clear that the values of the Keyfitz's index of entropy could be used for estimation of the maximally possible improvement of the life expectancy – value of the index multiplied by one hundred shows the possible percentual increase in the life expectancy, if there is a theoretical 100 % decrease of mortality (specific mortality rates). Of course, this is only a simplified model approach, because logically, the decline of the age-specific mortality rates by 100 % at all ages would cause these rates to be zero and the life expectancy would theoretically grow beyond any limits. However, for the needs of the analysis and quantification of the theoretically possible increase of the present values of life expectancy, we allow this simplification in this regard. Thus we can easily estimate the average maximal value of life expectancy (or "average maximum life span") that could be achievable under the condition of improvement of the existing level of mortality (Nagnur, 1986).

In the demographically most developed countries, the average maximum life span approaches values above 90 years (Figure 61). Nevertheless, in the Czech Republic the value is still lower in comparison with the most developer countries. This is due to a lower present value of life expectancy in the Czech Republic compared to these countries, because in this approach, the average maximum theoretically attainable life span is calculated as the percentage increase of the present value of life expectancy. The indicator therefore reflects the current situation in a given population.

Clearly the development could be divided into two phases. In the post-communist countries the stagnation from 1960s to the end of 1980s could be seen for females, for males it was rather a decreasing tendency (except for the Czech Republic). After the change of the political regime the studied post-communist countries could be divided into two groups – those where the situation started to improve (the Czech Republic, Poland, Hungary) ant those where the situation got even worse (Russia and temporary also Bulgaria; in both those countries the reason could be found in the economical development and transformation of the society). In Sweden (and this tendency could be visible also for more Western and Northern European countries) the development was significantly different for males and for females. For females the increasing tendency could be seen already from the World War II, for males it was started around the 1980s. Such a developmental pattern will be visible also in other analyses later in the work and we will try to explain it by the distinguishing of two components of mortality in the next chapter.

Figure 61: Average maximum life span derived from the Keyfitz's index of entropy, selected countries, by sex, 1950–2006



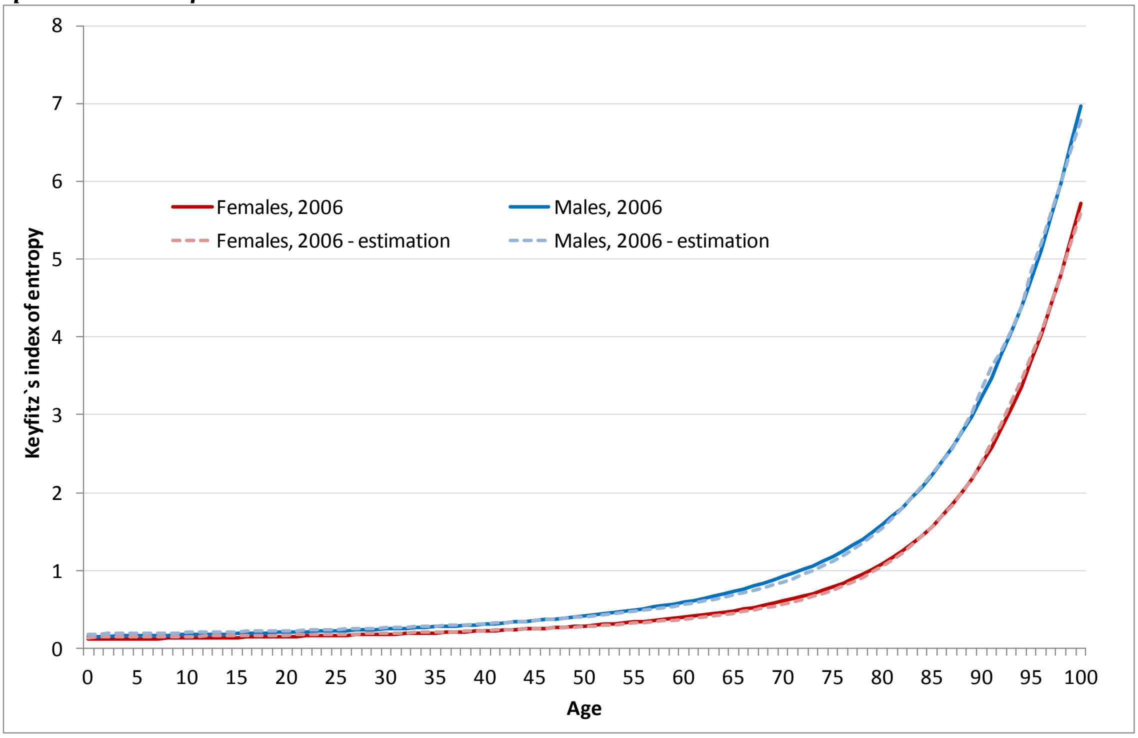
Source of data: author's calculation based on Human Mortality Database (2010)

From the values of the Keyfitz's index of entropy and from the life expectancy it is evident that both this indicators have partly similar pattern of dependence on age, but opposite in its trend. This led to attempts to find and express a specific model of the relationship between the life expectancy and the Keyfitz's index of entropy. Nagnur (1986) proposed the formula:

$$H_x = A + B * \left(\frac{1}{e_x}\right) + C * \left(\frac{1}{e_x}\right)^2$$

with parameters *A*, *B* and *C*. This model seems to be the best one among others possible expressions. Using the data of the Czech Republic in 2006 for both sexes, this equation really fits very well (see Figure 62), confirming that the Keyfitz's index of entropy is actually easy predictable from the life expectancy and thus the life expectancy could be well estimated from the values of the Keyfitz's index of entropy. The values of the estimated parameters in our model are partly different in comparison to results of Nagnur (1986) – in general, the parameter *A* reaches a very low values, the parameter *B* could be usually expressed in the range of units (it is lower than 10), and the parameter *C* was for the Czech data estimated as to be very similar to the values of the parameter *B*, in the study of Nagnur (1986) its published values were significantly lower (close to 1 or even negative). Except of evaluation of this relationship there are not many possibilities of its practical usage. The only one which could be mentioned is in demographic forecasts where it could be used for the assessment of the estimated life expectancy using the extrapolation of the Keyfitz's index of entropy.

Figure 62: Estimation of the values of the Keyfitz's index of entropy in dependence on the life expectancy according to age, comparison with empirical data, Czech Republic, by sex, 2006, the least squares method of estimation used



Source of data: author's calculation based on Human Mortality Database (2010)

7.3.2 Calculation of the interquartile range according to age, decomposition of the interquartile range, introduction of the concept of mortality shifting

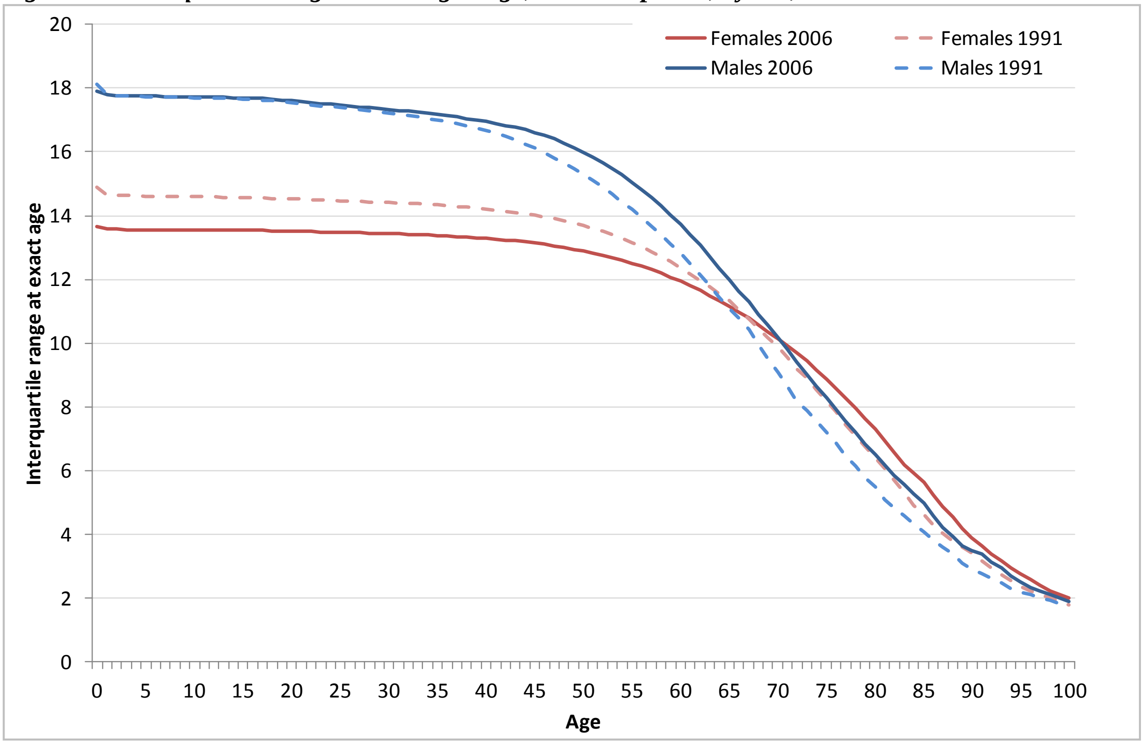
The advantage of the interquartile range is the possibility to work more in depth with this indicator. As was mentioned above, this indicator could be calculated not only for age 0 (or other selected age) but also for all ages in the life table. That means that the indicator is constructed in the same way for all the ages. In the calculation all the ages below the currently studied one are omitted.

Relevant demographic methods of decomposition could be also applied to this indicator. Thanks to the decomposition it could be revealed which ages stand behind the difference of its values between two studied populations. Thanks to those deeper analyses it is possible to learn more about the development of mortality as it is expressed by the indicator of interquartile range. This fact will contribute to the distinguishing between the process of rectangularization, shifting and derectangularization of the survival curve.

7.3.2.1. Calculation of the interquartile range according to age, distinguishing among the terms shifting, rectangularization and derectangularization

There occurred an interesting situation in terms of development of interquartile range according to age in case of females in the Czech Republic between 1991 and 2006 (Figure 63). Values of the indicator decreased during this 15 years for younger ages, i.e. that the total variability of length of life dropped; however, for higher ages (approximately over 70 years) the interquartile range has increased – so the variability of ages at death increased at higher ages.

Figure 63: Interquartile range according to age, Czech Republic, by sex, 1991 and 2006



Source of data: author's calculation based on Human Mortality Database (2010)

This indirectly confirms the hypothesis of the shift of the end of the survival curve to the higher ages. For males the interquartile range increased from a relatively early age – about 10–15 years. It could be said that the development of mortality rates during the last 15 years in the Czech Republic for younger ages rather confirms the rectangularization process; changes at higher ages indicate rather the derectangularization, if the rectangularization is understood as a process of decrease of the variability of ages at death (Fries, 1980).

The process of derectangularization, however, sounds like the opposite phenomenon to the rectangularization process. Therefore it could be also understood as opposite to the concentration of deaths at higher ages. In that case, this process would correspond rather to more uniform distribution of deaths in a wider age interval. The development of mortality rates seen for Czech females, however, suggests another phenomenon – as already mentioned – the shift of the end of the survival curve even more to the right (this shift can be described using the one-word term "shifting") (Canudas-Romo, 2008). Significant difference between these two terms, "shifting" and "derectangularization", can be illustrated again by the graph of interquartile range according to age for the Czech population (especially women) between 1991 and 2006 (Figure 63) – values of the indicator increased at older ages, but there is no evidence of any worsening of mortality conditions of older females (as might be assumed when using the term "derectangularization"), but again deaths are rather postponed to even higher ages. In this case, this development represents positive changes in mortality rates and it would be better to call the ongoing process as the shift of the end of the survival curve to higher ages ("shifting") than derectangularization (what could be understood rather as a synonym for a negative development).

This mentioned fact was already noted by demographers and it resulted to the concept of a so-called threshold age, which divides the age range into two intervals – to younger ages, where the postponement of deaths contributes to the process rectangularization, and to higher ages, where death postponement suggests rather the shift of the end of the survival curve to the right (Zhen, Vaupel, 2008).

As a conclusion of this brief sub-chapter all the three terms could be defined clearly because they will be used many times in the rest of this Thesis:

- For the purpose of this Thesis the process of rectangularization could be understood as such a development where the variability of ages at death is decreasing but the end of the survival curve does not change. If the modal age at death is rising then it is approaching to the maximal attainable age and as a consequence of that the standard deviation of ages at death above the mode is decreasing.
- Then the process of derectangularization could be described as such a development where the variability of ages at death is increasing as a consequence of rather negative trends. The modal age at death is decreasing and the end of the survival curve does not change or it is shifting to lower ages. As a reason of that also the standard deviation of ages at death above the mode is increasing.
- Finally the shifting process will be taken as a continuation of the rectangularization process when the variability of deaths stop to decrease and stabilize or it could slightly increase as a consequence of the shift of the end of the survival curve. The modal age is rising in this process. The end of the survival curve is shifted to higher

ages as the deaths are postponed. The result could be also the increase of the standard deviation of ages at death above the mode.

7.3.2.2. Decomposition of the interquartile range

The indicator of interquartile range enables also its deeper analysis using, among others, the decomposition methods (Weden, 2007; Wilmoth, Horiuchi, 1999). The difference of the interquartile range for two populations (assumed as involving all ages, i.e. calculated from the age 0) could be explained through decomposition methods, so the difference can be decomposed into contributions of particular age groups. One possible method was proposed and described for example by Wilmoth and Horiuchi (1999).

This method was applied by Weden (2007) on the population of the United States. She divided the population by age into five consecutive age groups of unequal length (0–4, 5–14, 15–44, 45–64, and 65 and more years), and calculated their contribution to the change of IQR_0 in seven 15-year periods. As her results show, rather younger ages are contributing to a reduction of interquartile range and on the contrary higher ages tend to increase values of the indicator in time. People older than the age of the third quartile will have no impact on the development of the indicator and its changes.

The general stepwise replacement algorithm (Andreev *et al.*, 2002) was used for the illustration of possible decomposition of the interquartile range within this Thesis. This method was used through the implemented VBA macro in MS Excel. The program was produced by Shkolnikov and Andreev (2010). The general stepwise replacement algorithm makes it possible to calculate the age decompositions for any quantities based on life table functions. The basic idea says that the change of the quantity 1 to quantity 2 corresponds to the change of the vectors of age specific mortality rates for the two populations. This transformation is estimated by a “stepwise replacement” of age specific mortality rates in the first vector by rates from the second one (*ibid.*).

We consider $M^1_{[x]}$ to be the vector of age-specific mortality rates for the first population, m^1_y, where for ages $y < x$ these mortality rates were changed for the rates from the second population, m^2_y. Then $M^1_{[x+1]}$ is the vector of age-specific mortality rates for the first population where for all ages $y < x + 1$ the rates were replaced by rates of the second population. Under these assumptions the difference of mortality rates at age $[x, x+1)$, m^1_x and m^2_x, caused the change of the overall studied quantity expressible in the form (Shkolnikov, Andreev, 2010):

$$\delta^{2\to 1}_{x|x+1} = E\left(M^1_{[x+1]}\right) - E\left(M^1_{[x]}\right),$$

where $E\left(M^1_{[x+1]}\right)$ is the studied quantity (in our case the interquartile range) calculated from the vector of mortality rates $M^1_{[x+1]}$. For decomposition of life expectancy at birth, results of this method are exactly the same as the results of other methods – for example proposed by Andreev in 1982 or Arriaga in 1984 (Andreev *et al.*, 2002).

Similarly the opposite replacement could be used in the calculation (Shkolnikov, Andreev, 2010):

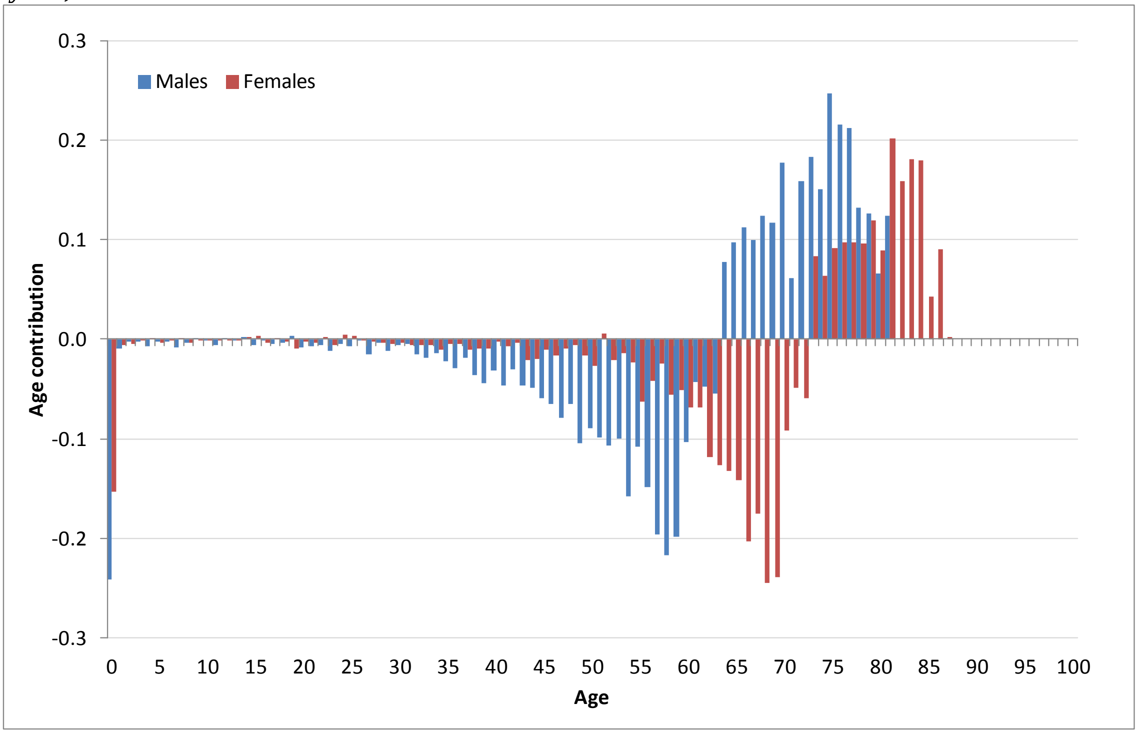
$$\delta_{x|x+1}^{1\to 2} = E\left(M_{[x+1]}^{2}\right) - E\left(M_{[x]}^{2}\right).$$

The final age-specific contributions are then calculated as (Shkolnikov, Andreev, 2010):

$$\delta_{x|x+1} = \frac{\left(\delta_{x|x+1}^{2\to 1} - \delta_{x|x+1}^{1\to 2}\right)}{2}.$$

Because this type of decomposition is an additive one, it must hold that the overall change of the studied quantity is a sum of the final age-specific contributions calculated in the previous step.

Figure 64: Age-specific contributions to the change of interquartile range (in years), Czech Republic, by sex, 1990–2000



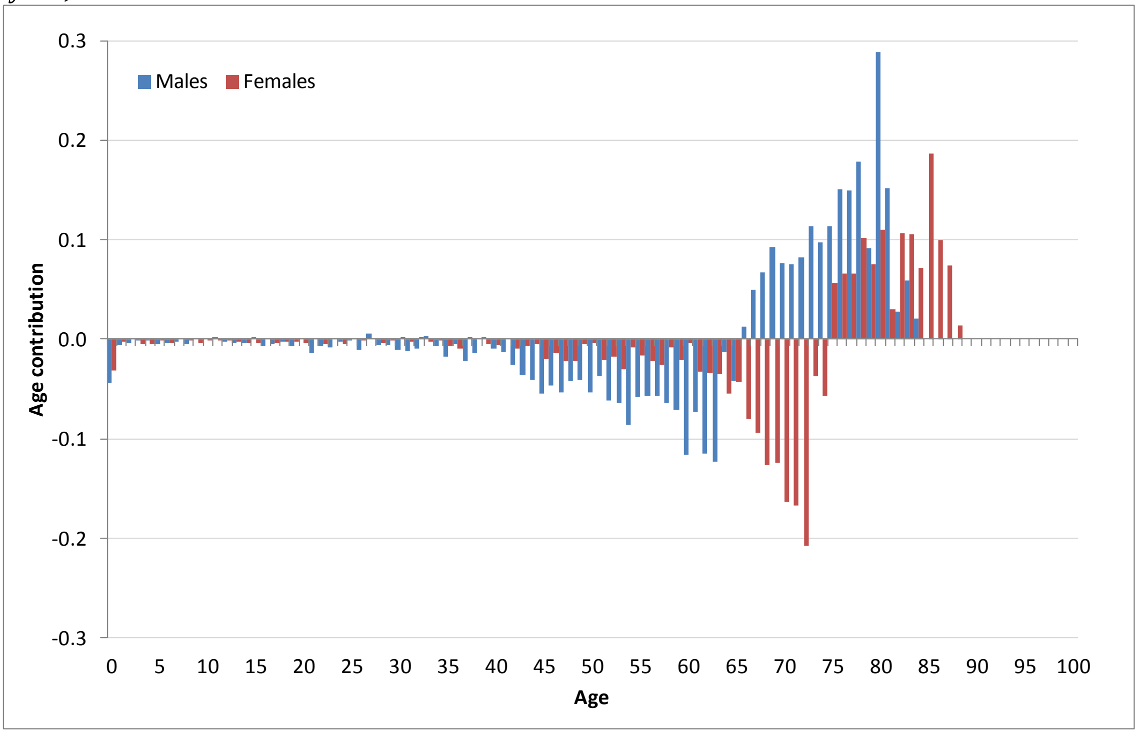
Source of data: author’s calculation based on Human Mortality Database (2010) and the VBA/Excel macro (Shkolnikov, Andreev, 2010)

The example of the Czech Republic (Figures 64 and 65) illustrates the contributions of males and females to the change of interquartile range between 1990 and 2000 and between 2000 and 2008. The results support findings of Weden (2007), although the calculation was not for age groups but for units of age and although she used a slightly different method of decomposition.

The basic results are almost the same. While at younger ages the end of the lower quartile is shifted to higher ages due to improvements in mortality rates (and thus the value of the interquartile range tends to be reduced), for the same reason there is a shift of the beginning of the upper quartile to even higher age. That causes “the pressure” on lengthening of the indicator. The sum of these negative and positive contributions of the individual age groups is then the overall change of the indicator value. Level of mortality at the highest quartile and its changes then has logically no effect.

Then it could be summarized that when the mortality improvements are more significant at rather younger ages then the indicator of interquartile range has the tendency to decrease. If the mortality improvements are more significant at higher ages (above ca 65 or 70 years for the Czech population) then the interquartile range has a tendency to increase. It is illustrated in the Figures 64 and 65.

Figure 65: Age-specific contributions to the change of interquartile range (in years), Czech Republic, by sex, 2000–2008



Source of data: author's calculation based on Human Mortality Database (2010) and the VBA/Excel macro (Shkolnikov, Andreev, 2010)

7.3.3 Relationship between the age-localization and the width of the interval of so-called "normal deaths"

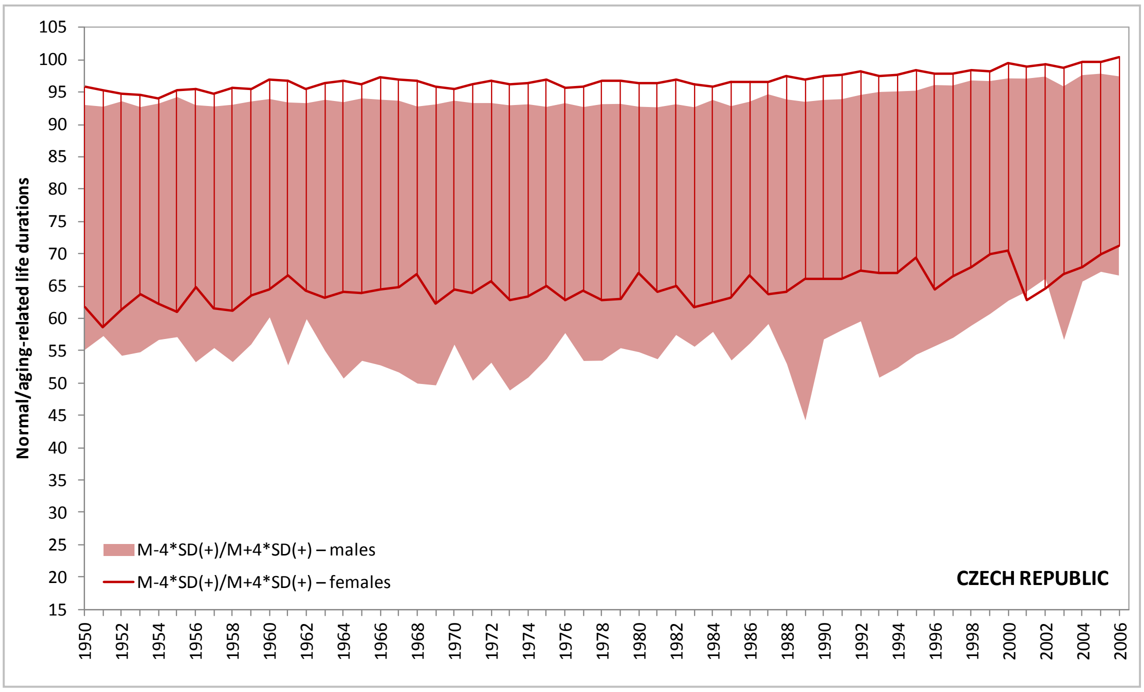
From values of the C-family indicators it is evident that the interval in which the deaths of a substantial part of the analyzed population occur is becoming shorter in time with mortality improvements. However, from these indicators it is not evident at what ages the considered interval lies (computation of these ages is obviously not impossible). To delineate these ages, where there is a substantial part of the deaths in the analyzed population concentrated, we can go back to Lexis and his definition of "normal deaths" (as an opposition to premature deaths), i.e. those which are result of aging ("age-related deaths") (Cheung *et al.*, 2005). There was a question for demographers how these "normal deaths" (or "ageing-related life durations") to define quantitatively. Kannisto (2001) suggested a possible approach to this definition. For simplification, he assumed an interval symmetric around the modal age at death. Then he expressed the range of this interval around the modal age as a *k*-multiple of standard deviation of ages at death above the mode and the constant *k* estimated as to be equal to 4 (Cheung *et al.*, 2005). In this way, it is possible to define the age range for deaths where it is possible to assume that those deaths are the so-called “normal deaths”, i.e. deaths directly resulting from senescent. This age range could be expressed as an interval equal to

$M \pm k * SD(M+),$

where M represents the modal age at death, $SD(M+)$ is the standard deviation of ages at death above the mode and k is the constant equal to 4 (*ibid*.).

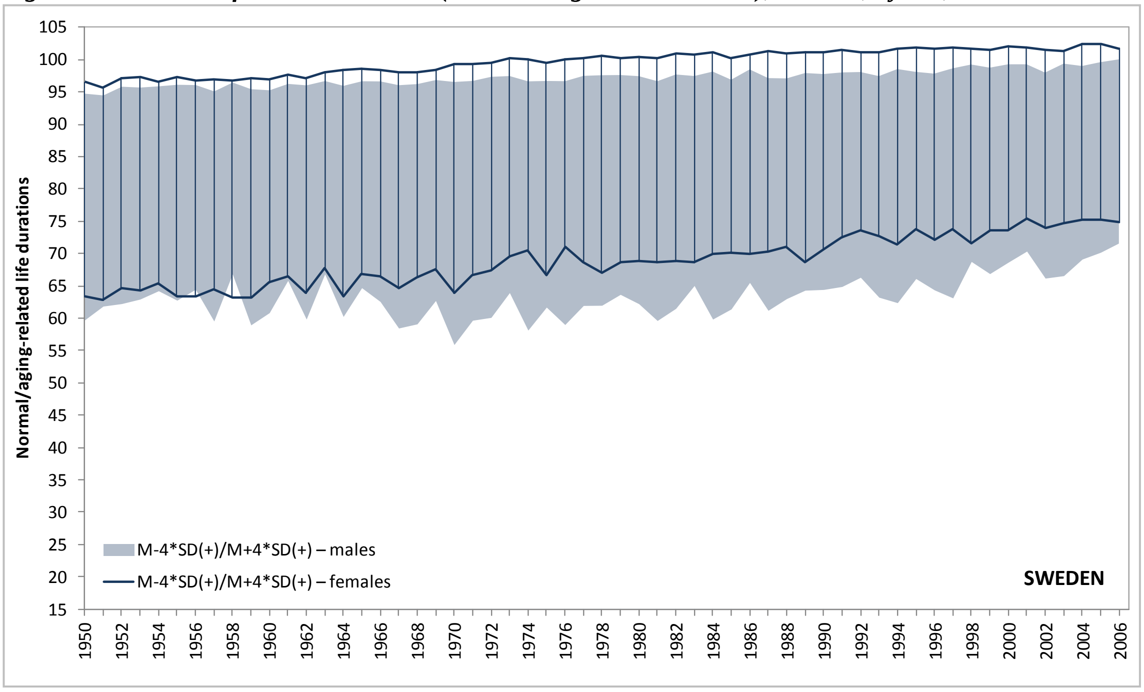
From the Figures 66 and 67 it is clear that the interval of ageing-related life durations becomes narrower in time (i.e. still more and more deaths are concentrated into a shorter age interval lying around the modal age of death) and moves to higher ages (indicating the assumption of the shift of the end of the survival curve). In the Czech Republic the interval is narrowing in general except from the temporary random fluctuations. The described trend was confirmed in the case of Sweden as well as the Czech Republic. In Sweden it seems that the interval is still shifting to higher ages although its width is constant. That would mean that the survival curve is shifting to higher ages while the variability of ages above the modal age at death is rather constant.

Figure 66: Interval of “normal deaths” (related to age and senescent), Czech Republic, by sex, 1950–2006



Source of data: author’s calculation based on Human Mortality Database (2010)

Figure 67: Interval of "normal deaths" (related to age and senescent), Sweden, by sex, 1950–2006

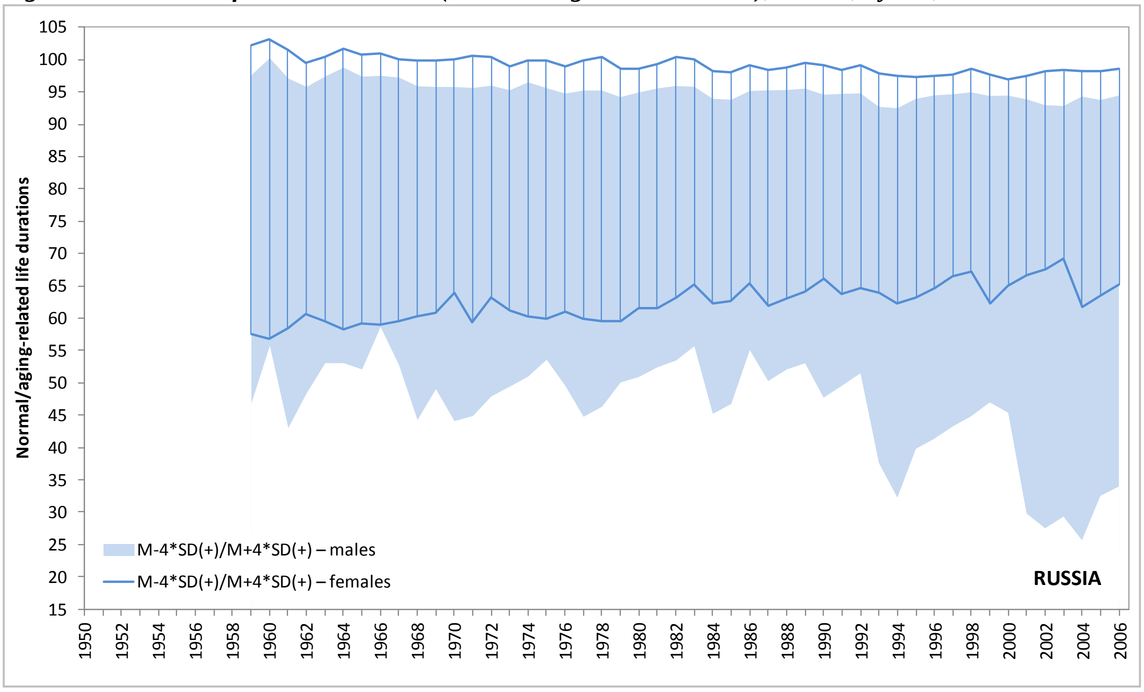


Source of data: author's calculation based on Human Mortality Database (2010)

Using the illustration of this interval it is also possible to show the specific development of mortality in Russia during the last decades (Figure 68). Especially noticeable is much longer interval (especially for males) than it was in the case of Sweden, but also in the case of the Czech Republic. This is the reason for significantly lower values of life expectancy and of the modal age at death of males in Russia in comparison to many other, also post-communist, countries. However, also the development of the length of the interval and its age position is easily observable. While for the Swedish and Czech data the interval became narrower in time (or at least the width does not change) and lay still at higher ages, in Russia for males the development is the opposite one. There the interval of "normal deaths" becomes on average rather extended and moves to lower ages. This is the confirmation of the impact of huge economical and social changes which negatively influenced mainly the mortality of adults at relatively low ages. Because the mortality development of males became significantly different from the development of females, it could be concluded that the reason for this is the different life style of males and females in Russia. In many studies the alcohol consumption is mentioned among others (Shkolnikov *et al.*, 2004; Kocová, 2009).

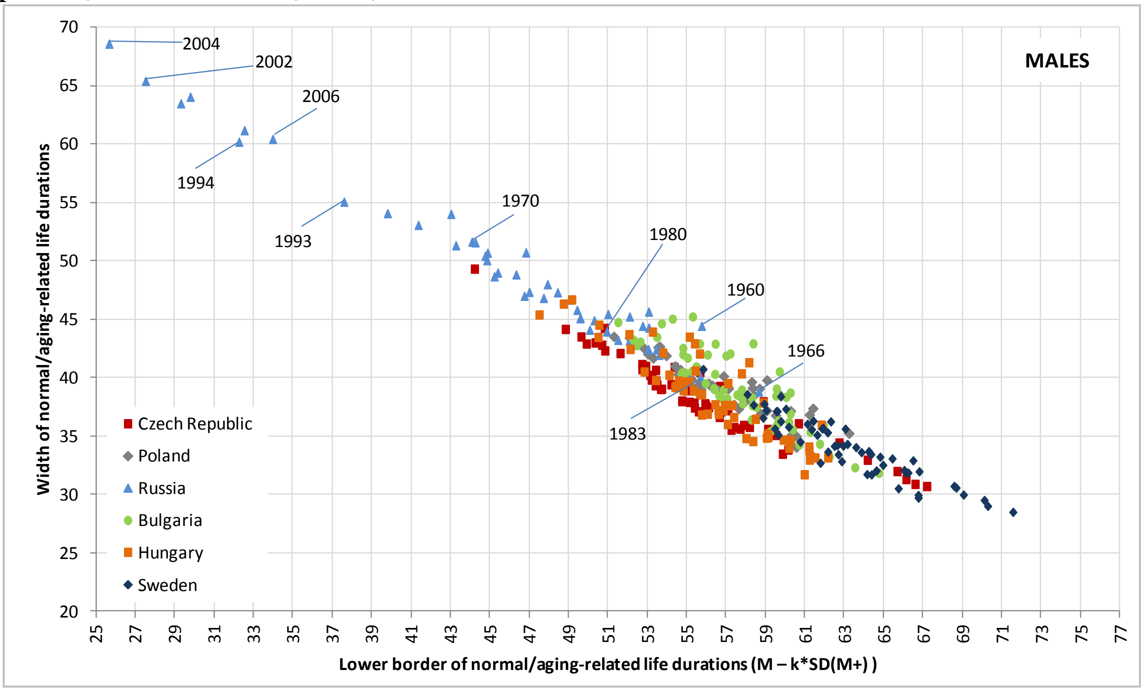
On the basis of the above presented results, an indirect dependence could be observed between the length of the interval of "normal deaths" and its position. This relationship is illustrated below, in the Figures 69 and 70, where the values in time have the general direction from the upper left corner toward the lower right one (with some random fluctuation of course).

Figure 68: Interval of "normal deaths" (related to age and senescent), Russia, by sex, 1959–2006



Source of data: author's calculation based on Human Mortality Database (2010)

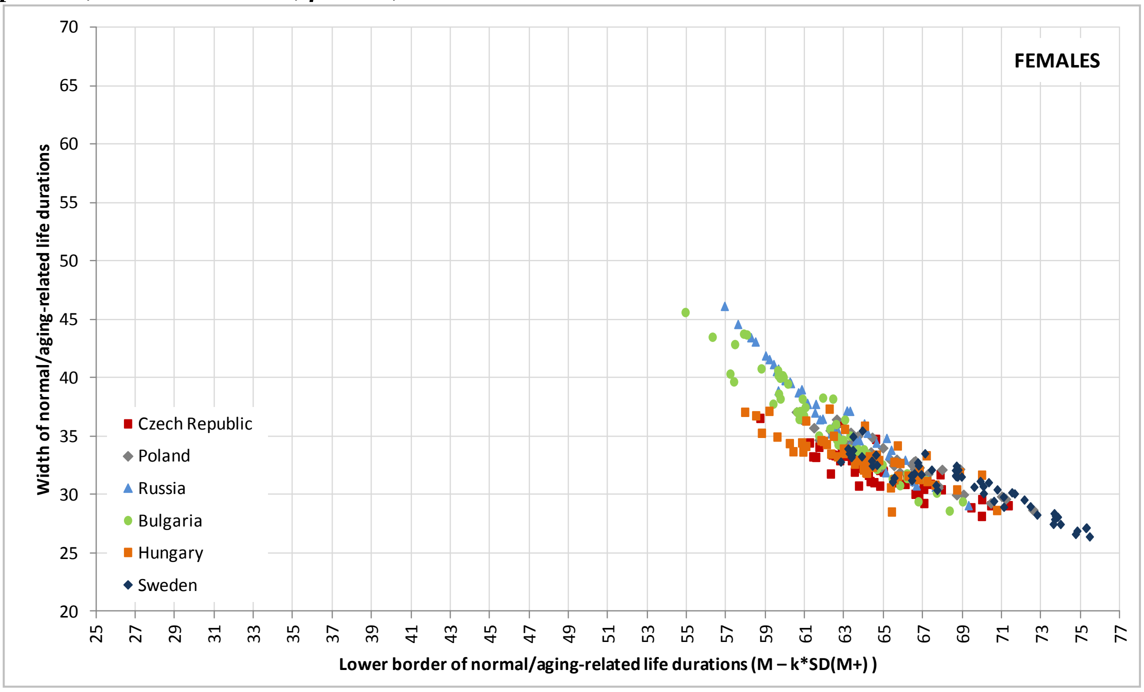
Figure 69: Indirect dependence between the length of the interval of "normal deaths" and its age-position, selected countries, males, 1950–2006



Note: years of selected values for Russia are marked

Source of data: author's calculation based on Human Mortality Database (2010)

Figure 70: Indirect dependence between the length of the interval of "normal deaths" and its age-position, selected countries, females, 1950–2006



Source of data: author's calculation based on Human Mortality Database (2010)

Again the only exception from the described rule is Russia, specifically the Russian males. For them, as already mentioned, the development is quite the opposite one. Based on the data used it is clear that in the postwar years the interval of "normal deaths" began at about the age of 50 years and lasted approximately 45 years. Still it holds that the quality of this data is certainly questionable. At the beginning of the 21st century the outset of the interval for males moved to ages about less than 30 years and it takes more than 60 years. So the direction of development (for Russian males only) can be illustrated in the graph as the movement from the lower right corner toward the upper left one. Values for selected years are marked in the graph (Figure 69). For females this exemption does not apply and the trend of development remains the same as in the other analyzed countries.

7.4 Rectangularization, derectangularization or shifting?

Based on the results for all the indicators presented above, it can be assumed that in all the countries selected and analyzed here there still dominates the development of mortality corresponding with the process of rectangularization and compression of mortality. With respect to the values of the indicators it can be stated that the survival curve on average still more and more resembles the shape of a rectangle and that the variability of ages at death decreases in time. Nevertheless, as mentioned in the text above, the development of values of many indicators in recent years is slowing down. It was visible for example for Swedish data, where Sweden was taken to the analysis as a representative of demographically and economically developed countries. This trend of stagnation is often not due to negative development of mortality in the analyzed countries (except of the specific case of Russia) which could be characterized by the increased intensity of mortality. Therefore it probably would not be appropriate to refer to this process as "derectangularization". In connection with the aforementioned facts it was formulated the hypothesis of the shift of the end of the survival curve, "the shifting mortality hypothesis" (Canudas-Romo, 2008). The hypothesis is based on the assumption that despite the decline in mortality the shape of the force of mortality curve does not have to change necessarily, but rather it can simply shift without any change of its pattern. As a consequence of that, also the variability of ages at death does not have to change although there is a change (increase) of the modal age (*ibid*.).

The development of the results of some of the indicators mentioned above is in favor of the shifting mortality hypothesis. This was evident for example in the case of the interval of "normal deaths" (or "ageing-related life durations"). The width of this interval in the developed countries with generally low level of mortality (e.g. Sweden, it is more evident in the case of females) changes still less and more slowly. It could be said that the values are rather stagnating at relatively low levels. Despite the fact that the width of this interval does not change significantly, the interval is still moving to higher ages. Hypotheses of compression and shifting of the end of the survival curve were studied also by Zhen and Vaupel (2008). They defined so-called threshold age, reduction in mortality at ages below this threshold age results in favor of the mortality compression. Improvement of mortality above this age has the opposite effect and thus leads rather to the concept of shifting of the survival curve to the right (*ibid*.).

The derectangularization process, if it is understood more in a negative meaning (greater variability in the ages at death involving also younger ages and rising deviation of the survival curve from the rectangular shape) reflects rather worsening of the overall mortality. During the studied period we have witnessed such a development only in Russia (from all the considered countries) where it was due to specific social, economic and other conditions (Gavrilova *et al.*, 2002; Shkolnikov *et al.*, 2004; Kocová, 2009).

If we do not assume a negative turn in rather positive development of mortality in demographically developed countries, in the future we can expect slowing of the rectangularization process or compression of mortality and further development corresponding more with the shifting mortality hypothesis. Because of the very background theory of the rectangularization process as a possible approach to revealing and proving the existence of a theoretical limit of human life span, one above mentioned assumption is confirmed. If such a limit exists, it cannot be proved or rejected on the basis of the results obtained for the indicators and rather it can be assumed that the human population by its current development in mortality is not approaching significantly to such a potential limit (Wilmoth, 1997).

7.5 Summary

The rectangularization process is directly linked to the process of mortality compression. Both processes basically represent the same, which is expressed by the shape of the survival curve approaching to the rectangle or by changing the shape of the curve of the distribution of table deaths toward greater concentration of deaths around the modal age at death (Fries, 1980). The hypothesis of rectangularization or compression is, however, only one of the three hypotheses that try to prove or disprove the existence of the limit of human life span (Wilmoth, 1997).

Alternatives to this hypothesis are for example the assumption of a fixed limit of human life or of limit distribution (*ibid*.). However, the concept of rectangularization of the survival curve or compression of mortality itself is so wide that in demography it has developed into a separate part of the mortality analysis.

Several of many possible approaches to the analysis of the rectangularization process were introduced and presented in the text. Some indicators that can be used for the description, are rather traditional in demographic analysis and also commonly used (for example, the modal age, interquartile range), or their slightly modified and supplemented versions could be used (such as standard deviation above the modal age at death, etc.). An individual group of indicators consists of those designed specifically for the analysis of the rectangularization process (e.g. indicators of the C-family) or derived from the indicators used in other fields of science (such as the Gini coefficient used primarily in economics or an index of entropy known from statistics).

Changes in mortality that occurred in the analyzed countries were presented through all the introduced indicators. Especially evident was the rapid improvement of mortality (mainly in the post-communist countries) after the World War II and the instability of development prior to this period (what was illustrated by the development in Sweden where data is available in a longer time series). In demographically developed countries there the change of the pattern of the distribution of table deaths is increasingly apparent. This change corresponds with less variability of ages at death and greater concentration of deaths around the modal age. It was also shown that further extension of human life and decrease in mortality at the highest ages could play a significant role in the future development of the process. Such positive changes can lead again to higher variability of ages at death.

The ongoing process of rectangularization, and thus also compression of mortality, helped to express the postponement of deaths to higher ages, where mortality became much more concentrated. The whole process thus reflected the improving mortality especially at young and middle age. At the present time when mortality improves at the highest age groups (generally above the modal age at death), the process of rectangularization slows down or even reverses. It is a consequence of postponement of deaths from the already high ages (above the modal age) to even higher ages. This can lead to stagnation or even re-growth of variability of ages at death. Such a process should be called “shifting of the survival curve” rather than “derectangularization”, because it is not the opposite process to rectangularization, but rather a process following to it (Yashin *et al.*, 2002). On the other hand, the term derectangularization could be undoubtedly used for the indication of the negative development for example in Russia at the time of the mortality crisis.

When it's over, I want to say all my life
I was a bride married to amazement.
I was the bridegroom, taking the world into my arms.

When it's over, I don't want to wonder
if I have made of my life something particular, and real.

I don't want to find myself sighing and frightened,
or full of argument.

I don't want to end up simply having visited this world.

Mary Oliver

Chapter 8

Background and senescent component of adult mortality, shifting hypothesis and the proposal of age-specific shifts

In the previous chapter, there the processes of rectangularization, derectangularization and mortality shifting were briefly defined. In this chapter the third one, the mortality shifting, will be studied in more detail. The motivation for this study has more reasons. First of all it showed out that the mortality shifting becomes the reality of mortality development above all in low-mortality countries. That means that the knowledge of this process could significantly help in demographic forecasting of mortality. Besides the mortality shifting is also used in other parts of mortality analysis (as will be described later, the process of shifting is the key assumption in the analysis of so-called tempo effects, etc.). At the beginning of this chapter the decomposition of mortality into a background and senescent component will be introduced. For an independent study of the senescent mortality component (i.e. the mortality component which is age-dependent) often the mortality shifting hypothesis is used (Bongaarts, 2005). Such an application of this concept is however conditioned by one important assumption, which will be verified later in this chapter. Based on the results the mortality shifting will be studied in more detail in practical as well as theoretical (or formal) point of view. The results of this chapter are then important in the following one where the mortality shifting will be in the role of a key assumption of the analysis of tempo effects, as was already mentioned.

8.1 Introduction and the aim of the chapter

Rapid and deep changes in the mortality development during the past decades motivated demographers to analyze the trends more deeply and to develop specific methods which could be useful not only for the description but also for forecasting of the adult mortality. Because of the already very low level of mortality at younger ages, the concentration of demographers started to be focused mainly on higher ages and the population group of the, so-called, "oldest-old". For accurate demographic forecasts of these age groups the analysis of mortality development at higher ages has become a key factor (Gavrilova, Gavrilov 2011).

Among others, methods of mortality analysis differ according to data which are used in them. When there the individual data are available, some types of survival models could be used (Aalen *et al.*, 2010) or the influence of selected risk factors could be tested in regression analysis (logistic regression, etc.; e.g. Rychtaříková, 2008). When there is a possibility to use cohort mortality data one can consider the usage of for example the frailty models which can handle the unobserved heterogeneity of population (see Chapter 10). These types of approaches become popular (among others) during the last years and are still developed rapidly and highly promising also for the future (see e.g. Vaupel *et al.*, 1979; Wienke 2011; in the Czech Republic Koudelka, Lustigová, 2010 or Hulíková Tesárková, 2012).

Very often it is not possible to use the cohort or individual data and the only possibility for a demographer is to analyze the trends in cross-sectional aggregated datasets. Also for this kind of data many types of analysis could be used. One of the basic tools of mortality analysis, the life table, is often applied to cross-sectional data (almost the whole first part of this Thesis was devoted to this issue) or some of many parametric, semi-parametric, or non-parametric functions could be used for the mortality analysis (several parametric methods were described in Chapter 5, non-parametric and semi-parametric methods are not the subject of this Thesis).

Within the analysis of the force of mortality (usually expressed in the form of some mortality law, i.e. parametric function) the decomposition of the total mortality into two different components (mortality component dependent on age and independent on age) is applied frequently (among other methods). Then both these components could be studied separately. The latest or most famous studies dealing with this topic (Bongaarts, 2005; Bongaarts, 2009; Gavrilova, Gavrilov, 2011; Gavrilov, Gavrilova, 1979; Gavrilov, Gavrilova, 1991, etc.) presented the distinguishing to the senescent (the term proposed by Bongaarts, 2005) and background mortality (Gavrilov, Gavrilova, 1979 used the term "base-line mortality").

Bongaarts based his analysis on the application of the logistic model (more details below) while Gavrilov and Gavrilova traditionally use the Gompertz-Makeham formula (more details below). Both groups of authors tried to apply this type of analysis to the data from developed mostly European countries and to distinguish the two mentioned components of mortality. It could be done also by the separate study of the development of particular parameters of the used function. Bongaarts (2005) concentrated mostly on latest decades of the 20th century (after the World War II), Gavrilova and Gavrilov (2011) focused on longer time interval – almost the whole 20th century – in Sweden and the USA.

In the Czech Republic a similar approach was applied already in the 1980s for example by Koschin (1989) who studied the development of particular parameters of the Gompertz-Makeham function in Czechoslovakia during the period 1960–1986 which were compared with the development in Sweden. From the developmental patterns of the parameters the possible future trends were estimated.

This chapter will follow the issue of senescent and background mortality as was stated by the authors mentioned above and extend the analysis to wider range of countries, also to the post-communist countries. In the following text, these two approaches (the decomposition on background and senescent mortality applied to logistic as well as the Gompertz-Makeham model) will be introduced briefly (the introduction will be based mainly on previously cited authors) and then applied to the data from not only the developed Western European countries. Also the study of development of particular parameters in time will be used. Moreover, where possible, longer time interval will be considered in comparison with previously published works (cited above).

In the second part of the chapter the shifting mortality hypothesis (mentioned already in the Chapter 7) will be introduced in more detail and described from various points of view (Canudas-Romo, 2008; Kannisto, 1996; Bongaarts, 2005, etc.). This hypothesis is closely related to the previously mentioned decomposition of the total mortality and describes the shifts of the mortality curve or separately only of the senescent mortality curve (for its connection with the concept of rectangularization of the survival curve see Chapter 7). Special approach within the shifting mortality theory is the study of the movements (shifts) of the life table functions when depicted in the graph. In the last part of the chapter some theoretical relations dealing with the assumption of the mortality shifting will be developed.

Based on the introduction to this chapter, its main goals could be summarized as follows (the goals are then reflected also in the structure of the chapter):

1) The first aim of the chapter is to find out whether there are some differences between the results for two groups of countries (non-post-communist, labeled as "NPC countries", and post-communist, labeled as "PC countries"), and whether the considered approach to the decomposition of the mortality to the senescent and background components reveals some more universal pattern of mortality development in both groups of states. The analysis will be focused more on the Eastern European countries (in comparison with the authors cited above) and it's comparison with Western and Northern Europe. The focus will be set on the analysis according to sex, development in time and age differences. For the estimation of parameters the method of weighted non-linear least squares was used, i.e. a slightly different method in comparison to the previously mentioned papers (Bongaarts, 2005; Gavrilova, Gavrilov, 2011).
2) The second aim of the chapter was to apply the concept of mortality shifting hypothesis to the results of the previous analysis and to study the mortality shifts. In this phase of the work the conditions set by Bongaarts (2005) will be partly left behind and the shifting process will not be analyzed under the condition of a parallel shift of the whole hazard (or survival) curve (as stated by Bongaarts, 2005 and showed not to be valid for less developed or historical populations, see later within this chapter). Finally there will be proposed a simple method of the estimation of "age-specific shifts" and the overall condition under which the process could be noted as "mortality shifting" will be derived based on this estimation.

3) For closing of the theme and logical connection of all its aspects, the attention will be given in a theoretical way also to formal relationships within the life table under the condition that there is the process of parallel change of mortality rates in it. Thanks to that the general conditions for the mortality shifting will be developed.

The research questions involved in the above mentioned goals of the chapter representing the partial aims of the analysis can be then summarized as follows:

1) Are the results of Gavrilov, Gavrilova (1979; 1991; 2011) and Bongaarts (2005) valid also when using different method of parameters estimation of the models (weighted non-linear least squares)?
2) The above cited papers show the results mainly for Sweden, the USA and other mostly Western European countries. The question is whether the development is similar also in other countries considered in the analysis within this Thesis.
3) What are the differences between the post-communist and the other developed countries?
4) What are the differences of the results of the decomposition to senescent and background mortality according to sex?
5) What are the differences of the development of the senescent mortality component according to age?
6) Is there a parallel shift in time of the mortality curve to higher ages or the shifts are significantly different according to age?
7) According to the results of the previous point – when there could be indentified the shifting process in the data (in the sense of a parallel shift)?
8) How the life table functions respond to the change of the mortality curve in general?

The questions 1 to 4 will be solved in the first part of this chapter, where the decomposition of mortality into two basic components will be presented. The questions 5 to 7 are related to the second part of the chapter dealing with the shifting mortality process and its identification in data. The last question became the topic of the third part devoted to formal expressions of the relationships within a life table under the condition of the parallel change of the mortality curve.

Analyses in this chapter ale based mostly on the usage of the data from the Human Mortality Database (http://www.mortality.org)[13]. For the analysis almost all the countries with available data were used[14].

[13] The only exception is the Czech Republic, where the time series used in this chapter is longer than the time series published in the Human Mortality Database. The used data series starts at 1920 (but data are not available for years 1938–1944). The source is the Czech Statistical Office. (http://www.czso.cz/csu/redakce.nsf/i/umrtnostni_tabulky_za_cr_od_roku_1920/$File/cr_ut_1920_2010.zip).

[14] The analysis was made for Australia, Austria, Belgium, Bulgaria, Belarus, Canada, the Czech Republic, Germany, Denmark, Spain, Estonia, Finland, France (total population), the United Kingdom (total population), Hungary, Switzerland, Chile, Ireland, Iceland, Israel, Italy, Japan, Lithuania, Luxembourg, Latvia, the Netherlands, Norway, Poland, Portugal, Russia, Slovakia, Slovenia, Sweden, Ukraine, and the USA.

8.2 Theoretical background

In fact the original idea of the decomposition of mortality into a senescent and background mortality could be followed already to Makeham and his proposal of the constant which enriched the Gompertz formula (Makeham, 1860). This constant could be taken as a representative of such mortality which does not depend on age. The variable term in the equation represents the increase of mortality with age as a result of the aging process (Gavrilov, Gavrilova, 1979). Since that moment demographers took the idea of two basic components as reasonable. For many years the Gompertz-Makeham formula was taken almost as a universal mortality law (Burcin *et al*., 2009).

Gavrilov and Gavrilova (1979) focused their attention on the long-term trends of the parameters of the Gompertz-Makeham function and used this approach for the analysis of age dynamics of mortality. On real data they proved, that the background mortality (they called it "base-line mortality") decreased in time. They showed that the parameter representing the base-line mortality was the only parameter of the Gompertz-Makeham function which changed in the analysis for the period 1911–1975 where the data for Sweden were used. It was presented in their work that the value of this parameter already after the middle of the 20th century decreased to a limit level close to zero. The authors then supposed that it was the reason for exhausting the reserves of future lengthening of the life span. On the other hand the senescent (age dependent) mortality component was nearly stable during the studied time period (Gavrilov, Gavrilova, 1979).

The same authors in one of their latest articles repeated their calculations. They involved a wider range of countries to the analysis (Gavrilova, Gavrilov, 2011). They noticed that the trend of development during the second half of the 20th century was different in comparison to their previous results. The background component was already at a very low level and it was almost stable in time, while the senescent mortality started to decrease in time. They found this new developmental trend to be the reason for the mortality decrease in the second half of the 20th century occurred in demographically developed countries with low mortality level (*ibid*.).

Bongaarts (2005) used a logistic mortality law formula for a similar analysis of the trends of the parameters as did Gavrilov and Gavrilova (cited above). In his work he analyzed a group of countries with low mortality level during the second half of the 20th century. He noticed that the slope parameter in the senescent component of mortality (see the formula later in this chapter) was nearly constant for all the countries involved in the analysis. Based on this assumption he concluded that the senescent mortality curve is shifted to higher ages without any change of its shape. Upon this fact, Bongaarts based the new logistic model usable also for forecasting which he calls the "shifting logistic model" (Bongaarts, 2005). He supposed the shift of the senescent mortality curve to be parallel according to age (what is based on the assumption of constant value of the slope parameter) and he also proposed the way how to estimate the amount of that shift (will be explained below in more detail).

In later works Bongaarts proposed an alternative way of the estimation of the background and senescent mortality components – he used the death rates according to causes of death (Bongaarts, 2009). Then he compared the basic characteristics calculated on the basis of the senescent mortality and on regular basis of the total mortality. Again it was proved that in the analyzed countries (17 developed low mortality countries) the "shape of the distribution of senescent deaths by age remains relatively invariant while the entire distribution shifts over time to higher ages as longevity rises." (Bongaarts, 2009, p. 203).

Both the groups of authors mentioned above, Gavrilov and Gavrilova and also Bongaarts, focused in their works on the adult mortality. The same will hold in this whole chapter. Mortality at lower ages will be neglected so that it cannot influence the results.

The decomposition of the total mortality into the component invariant with age and the component dependent on age has many important consequences for the further demographic analyses of mortality trends. First of all the pure trends of those both components could be studied, analyzed and then used in the forecasts (as discussed earlier).

Another way of usage of the results is connected with the shifting mortality hypothesis. As was also mentioned above, Bongaarts (2005; 2009) described the trend of the development of the senescent mortality as shifting to higher ages. This concept was already briefly mentioned in the previous chapter and it will be studied in more detail later.

The shifting mortality hypothesis became an almost independent issue in the mortality analysis. The concept was presented for the first time by Kannisto (1996). He described the process as a general move of the mortality curve to higher ages. He also suggested measuring of this shift in ages, as a difference of the two ages at which the same level of mortality was reached in two different moments in time (*ibid.*). As was already mentioned, Bongaarts (2005) used only the senescent component of mortality and proposed a method how to estimate the amount of the shift between the two defined moments in time. In this estimation he used the same definition of the shifting as Kannisto did.

Canudas-Romo (2008) described the shifting mortality scenario as "…a process that might be expected if the current mortality changes maintain their pace." (Canudas-Romo, 2008, p. 1179). He based his study on the modal age at death because he assumed the shifting as a development where the modal age rises, but the number of deaths at the modal age and the standard deviation above the mode seem to approach some limit value. He used several models of mortality change for the proof of his statement (Canudas-Romo, 2008).

Some revision of the studies dealing with the topic of the shifting mortality offers Zureick, who defines the shifting mortality as a shift in mortality rates across age (Zureick, 2010). Like Canudas-Romo, also she focused on the relationship with another demographic concept, the compression of mortality. She tries to prove (while using several simulations) that there could be found several conditions which stand behind the process of mortality compression, expansion, or shifting. She also pointed out that although the shifting could be taken as a movement of the mortality curve (understood as the curve of age-specific mortality rates) very often it is (and its consequences are) measured by the distribution of deaths and its variability (*ibid.*). She used several measures of the variability of ages at death (used usually in the analysis of the rectangularization process) and illustrated their development under several different conditions of the mortality shifting. It was proved that even in the case of parallel shift of the mortality curves across age there could be no mortality shifting detected when using the measures of variability of ages at death (or generally measures of mortality compression). In this chapter these results will be partially validated.

The third sphere of application of the results of the decomposition into senescent and background mortality component, which is however more connected with the previous section devoted to the mortality shifting, is the analysis of the tempo effect. The relation between the shifting mortality and the analysis of timing was introduced mostly by Bongaarts and Feeney (2002; 2003) and Feeney (2006). The Feeney's work seems to be the most relevant in relation to the shifting mortality concept and the estimation of the values of the shift of the mortality curve. Upon the survival curve he proposed the way how to estimate the horizontal differences between this curve in two different moments in time, i.e. so-called "the increments to life" (Feeney, 2006). The tempo effect, its assumptions and its role in the demographic analysis of mortality will be presented later in the text of this Thesis (see Chapter 9).

It could be seen in the literature that different authors define the mortality shifting in slightly different ways. Most of them (Kannisto, 1996; Bongaarts, 2005; Canudas-Romo, 2008 or Zureick, 2010) used the hazard curve for the definition, or the values of the age-specific mortality rates. They define the shifting mortality as a parallel horizontal movement of this curve. On the other side, there are some demographers who define the shift through the survival function (for example Feeney (2006); shifts of the survival curve are studied also in the rectangularization process, see Chapter 7). In the second part of this chapter, there we will consider the shifting mortality as the change of the hazard curve and estimate the age-specific shifts of the survival curve (similar to the Feeney's concept of increments to life) between two moments in time for selected countries. We assume that the real movement of the empirical curve is not parallel over age. Then the shifting could be defined according to these age-specific shifts as the process of the decrease (or eventually increase) of mortality when the age-specific shifts are constant (or nearly constant) across the age.

Finally, in accordance to the previously cited works, we will theoretically develop the relationships between the absolute or relative change (decline) in age-specific mortality rates and changes of other selected life table functions. It will be shown that when there is an absolute parallel decline of all mortality rates in the life table (i.e. all the age-specific mortality rates will decrease for some constant value), the relative change of the survival curve would be linearly increasing with age, the same is true for the distribution of deaths in the table (i.e. the relative change of these functions would be bigger at higher ages than at younger ages). When there is a relative parallel decrease of the age-specific mortality rates (i.e. the absolute value of the decrease is not the same for all the age-specific death rates but the relative decrease would be the same for all the age-specific death rates), the relative change of the survival curve and the curve of the table deaths will be nearly exponentially increasing with age, what causes that the absolute change of the values of this table functions is closer to the horizontal shift (when taking only the adult ages into consideration again). Generally it would mean that the mortality shifting (horizontal shift of the survival curve or density of deaths) could be observed under the condition that the percentage decrease of age-specific mortality rates is age-invariant. The proof and derivation of these relations could be seen below (Sub-chapter 8.5).

8.3 Mortality decomposition into the senescent and background components

8.3.1 Two components of mortality, two basic approaches

For all the countries involved into the study the decomposition of mortality into the background and senescent component was done. Two different approaches were used – the exponential and the logistic law of mortality. The first one was represented by the Gompertz-Makeham formula (Makeham, 1860) of mortality pattern and the other was represented by the Thatcher logistic formula (Thatcher, 1999).

Thus in the first case, the total mortality ($\mu_{x,t}$) at age *x* and time *t* was considered in the form:

$$\mu_{x,t} = A_t + B_t * \exp(C_t * x),$$

where *A*, *B* and *C* are parameters which can change in time *t*, and *x* is age. The lower indexes of the parameters signify the variability in time, usually they are not written in the standard notation of the equation. In the Czech demographic literature also a slightly different form of the Gompertz-Makeham function could be found (e.g. Koschin, 1989; Koschin, 2002; Pavlík *et al.*, 1986):

$$\mu_{x,t} = A_t + B_t * c_t^x,$$

but this form could be simply transformed to the form mentioned above:

$$\mu_{x,t} = A_t + B_t * c_t^x = A_t + B_t * e^{k_t^x} = A_t + B_t * e^{k_t * x} = A_t + B_t * \exp(k_t * x),$$

where $e^{k_t^x} = c_t^x$, i.e. $e^{k_t} = c_t$, k_t is a parameter changing in time but constant for all ages during one year (in the previous form of the equation marked as C_t), so as is the parameter c_t. Both these parameters are the same (only one of them is in an exponential form while the other is not). That is the reason why both the forms of the equation are equivalent.

The Gompertz-Makeham function could be divided (according to Gavrilova, Gavrilov, 2011, p. 110 with slightly different symbolization of the parameters) into the senescent component ($\mu_{x,t}^S$):

$$\mu_{x,t}^S = B_t * \exp(C_t * x)$$

and the background component ($\mu_{x,t}^B$):

$$\mu_{x,t}^B = A_t.$$

It is obvious that the background component of mortality does not depend on age and the senescent component is increasing with age exponentially. The sum of the senescent and background mortality is the total mortality.

In the second case the logistic model of the total mortality ($\mu_{x,t}$) was used in the form (Thatcher, 1999; Bongaarts, 2005, here again with slightly modified symbolization):

$$\mu_{x,t} = \frac{A_t * \exp(B_t * x)}{1 + A_t * \exp(B_t * x)} + C_t.$$

Again it can be divided into the senescent mortality ($\mu_{x,t}^S$):

$$\mu_{x,t}^S = \frac{A_t * \exp(B_t * x)}{1 + A_t * \exp(B_t * x)}$$

and the background component ($\mu_{x,t}^B$):

$$\mu_{x,t}^B = C_t.$$

And again it is clear that the background mortality does not change with age (but can change over time), while the senescent mortality is a function of age, and that the sum of both these components is the total mortality.

8.3.2 Methodology of the parameter estimation of the applied models

Very important part of the analysis is the estimation of the three unknown parameters of both the applied models (Gompertz-Makeham and Thatcher function). In comparison with many other demographic articles dealing with the topic of mortality laws (which usually used the method of the ordinary least squares or non-weighted generalized least squares), in this Thesis the estimation process was based on the weighted generalized (non-linear) least squares method with the usage of the statistical software SAS, version 9.2 (for the estimation the Gauss-Newton method was applied). The weights were used in order to eliminate the effect of heteroskedaticity of data, because the variability of mortality rates changes with age. The method of estimation was described in the Chapter 5, here it will be only briefly summarized. In practice again the macro introduced in the 5th chapter was used for the calculation.

The weights in the estimation method are considered to be reciprocal values of variance of the empirical data. The variance of the age-specific mortality rates could be written (according to Gerylová, Holčík, 1988) as:

$$\frac{m_x * (1 - m_x)}{P_x},$$

where m_x is the mortality rate at age x and P_x is the population alive in the middle of the studied time interval (year) at completed age x. The method of estimation is an iterative one; the weights are recalculated in each iteration of the procedure.

Other thing, which should be for clearness reminded here again, is the considered relation between the mortality rate (m_x) and the force of mortality (intensity of mortality μ_x). It is supposed that for all *x*, except the age zero, it holds (Thatcher, 1999):

$$\mu_{x+1/2} \approx m_x \ .$$

This assumption will be used in the estimation of the parameters.

The last question is how to select the age interval where the empirical mortality rates should be used for the estimation of the parameter values. In the paper of Bongaarts (2005) the ages above 25 years were considered. In this Thesis the same age was taken as the lower border of the age-interval so that the results are comparable. The age-interval used for the estimation in this work was limited by a maximal age of 90 – above this age the empirical mortality rates could be considered as less reliable and show usually high variability, on the other hand the aim was to use the most wide age interval as possible.

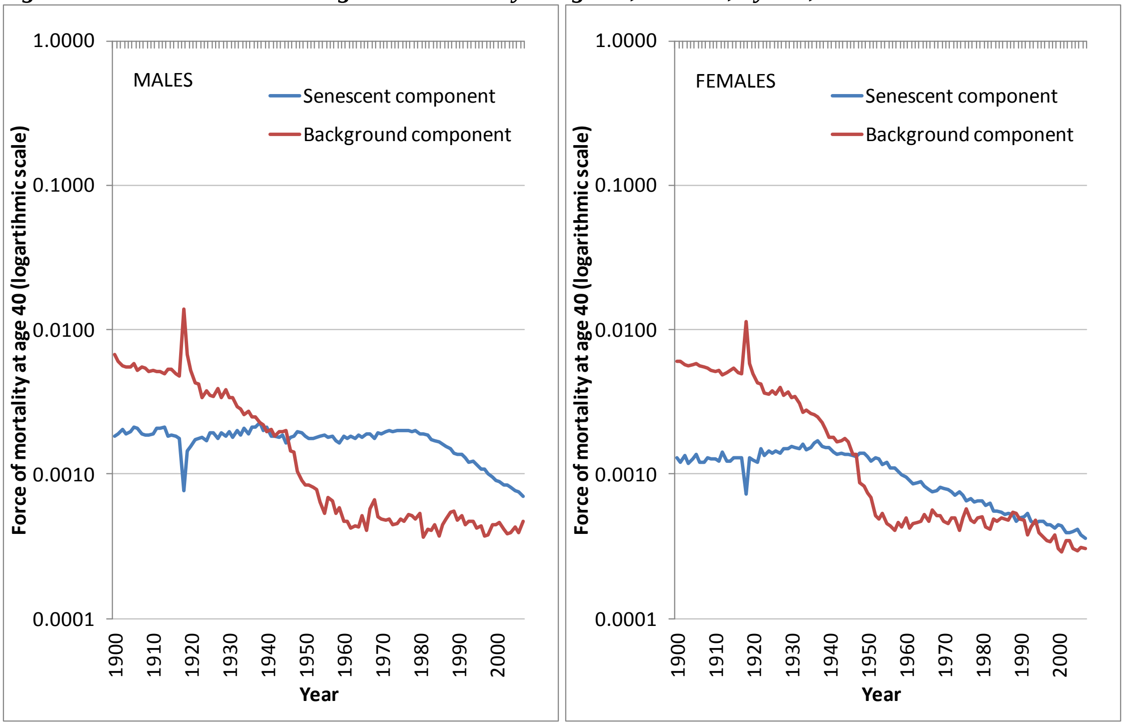
8.3.3 Basic results of the decomposition

Model 1 – the Gompertz-Makeham mortality law

Gavrilova and Gavrilov (2011) noticed that in the past the background component was decreasing with time (the rate of decrease was higher after the World War II) until it reached some low value which could have been considered almost as the limit value (Gavrilov, Gavrilova, 1979). This limit was reached during the 1950s by Swedish females and around 1965 by males in the same country (Gavrilova, Gavrilov, 2011). On the other hand the senescent component was almost constant in the past. This pattern was also verified by the further analyses applied to more data from various countries (Gavrilov, Gavrilova, 1991). But in later studies (Gavrilova, Gavrilov, 2011) they found that the senescent mortality started to decrease after 1950 for females and around 1980 for males (again on the example of Swedish data). These results will be confirmed also later in this work (although a slightly different method, presented above, of parameter estimation was used). At first the senescent mortality at age 40 was used (in accordance with the Gavrilova, Gavrilov, 2011) for the illustration of comparable results.

Through the first results (presented in the Figure 71) the conclusions of Gavrilov and Gavrilova (1979; 1991; 2011) were verified – by the estimation of the Gompertz-Makeham parameters it was shown that for the population of Sweden the total intensity of mortality could be derived into two components: the age-dependent senescent and age-independent background mortality. Also the second fact described by Gavrilov and Gavrilova (1979; 1991; 2011), that during the time the senescent mortality stayed nearly constant and started to decrease around the middle of the 20th century for females and in the 1980s for males, was verified for Swedish data and age 40 (age 40 was selected in accordance to the cited articles of Gavrilov and Gavrilova). Different development is confirmed, as expected, for the background mortality which was responsible for the total mortality decline before the middle of the 20th century.

Figure 71: Senescent and background mortality at age 40, Sweden, by sex, 1900–2008



Source of data: author's calculation based on Human Mortality Database (2010)

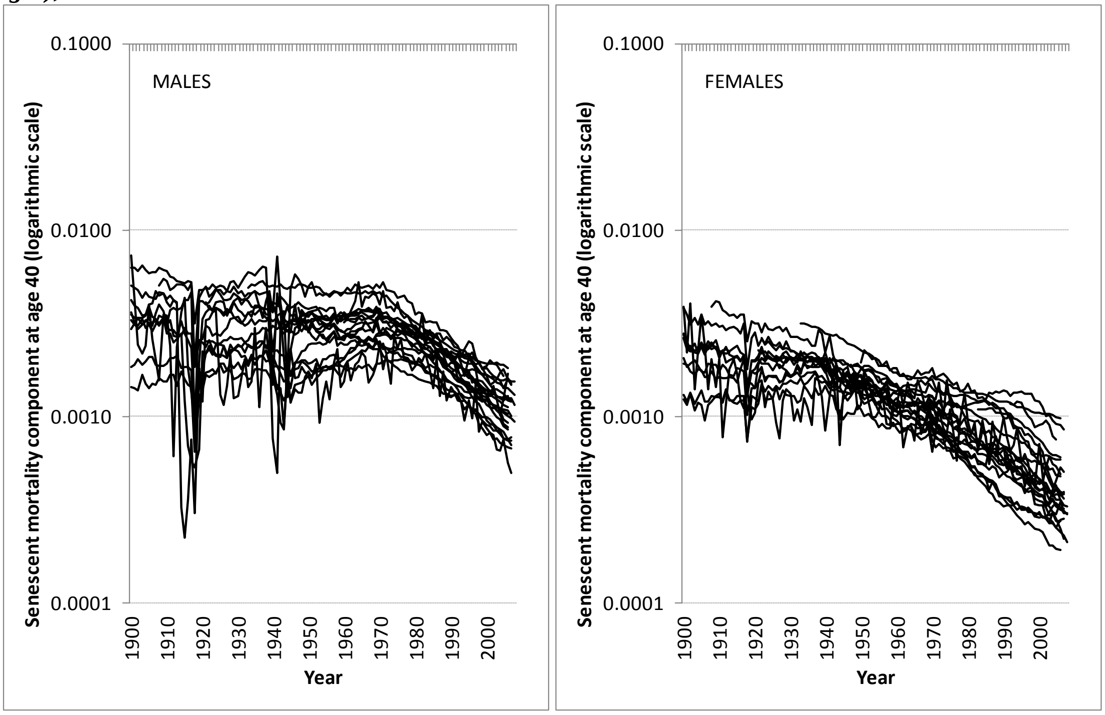
In the following part of the text, there the analysis goes more in detail. The development of the senescent and background mortality component will be studied in different countries, at different ages and also the comparison according to gender will be presented. For the ongoing analyses, two groups of countries were distinguished – the post-communist countries (labeled as "PC countries") and the non-post-communist countries (mainly the Western, Northern and Southern European countries together with the USA, Japan, Canada, etc., labeled as "NPC countries").

Regional differences

The aim of the following analysis is to find whether there are significant differences between the post-communist countries in the Central and Eastern Europe and the developed mostly Western and Northern European countries according to the development of the senescent and background mortality component in the past. In the following graphs the individual curves are not distinguished according to the countries because the aim is to study the development of both the defined groups of states as a whole and moreover the description of the symbols would be very complicated when there are so many curves in one picture. Some specific cases will be described in the text where needed but generally both the groups of countries will be considered as a whole and the average trends will be compared.

According to the distinguishing between Eastern Europe (considered here together with the post-communist countries in the Central Europe, "PC countries") and "other" countries ("NPC countries") we can see clearly the differences (Figure 72). Again we consider here the senescent mortality at age 40 as an example (comparable with the previous results). Generally males in the NPC countries (non-Eastern and non-Central European) have similar pattern as Swedish males do. The senescent mortality was almost constant during the past, except the periods of the World Wars. In all of the analyzed countries it started to decline between the end of 1960s and 1980s.

Figure 72: The senescent component of mortality at age 40, NPC countries, males (left), females (right), 1900–2008



Source of data: author’s calculation based on Human Mortality Database (2010)

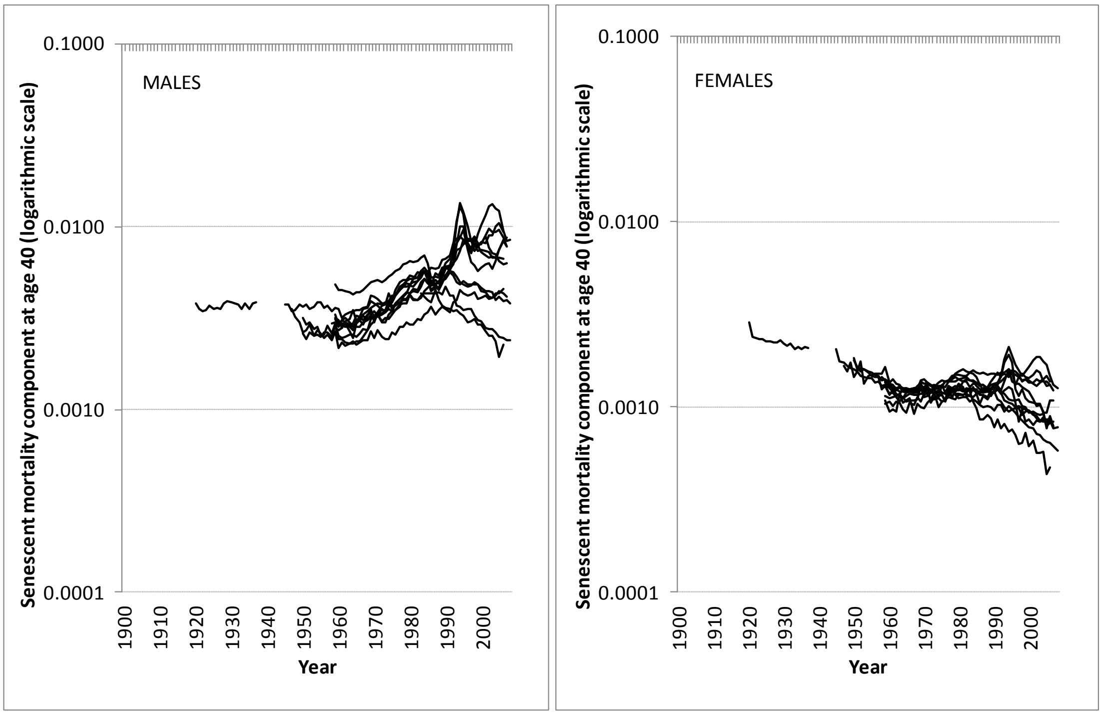
For females the development is slightly different – without any clear pattern which can be seen as a universal one. In some countries the trend of the senescent mortality at age 40 is decreasing already from the World Wars (Spain, Switzerland, the USA), in some countries the decrease started only one or two decades ago (Denmark). Timing of the decrease of the senescent mortality is the reason of different mortality development in time in various countries because, as will be shown later, the development of the background mortality is similar in all the NPC countries, at least for females.

The development is completely different in countries labeled as PC countries (or Eastern European; see Figure 73). For males the senescent mortality started to decrease only in some of the analyzed countries – the Czech Republic, Slovenia (during the end of 1980s) or in Slovakia and Poland at the beginning of the 1990s. In other countries, there the decreasing trend could not be seen and the trend is clearly increasing in time. The highest increase could be followed in Russia, Belarus or Ukraine. During the last decade the stagnation of the senescent mortality in those countries could be distinguished. During the 1960s to 1980s, when stagnation of the senescent mortality of males was shown in many NPC countries, there is a worsening trend visible in all the analyzed PC countries. It is likely to be the reflection of the social and economical situation in those countries which was also decreasing at that time (Holman, 2002; Shkolnikov *et al.*, 2004).

For females, the decreasing trend is apparent again only in some countries, like the Czech Republic or Slovenia, where it started during the 1980s and at the beginning of 1990s. In countries like Russia, Belarus or Ukraine almost no decreasing trend could be traced in the data.

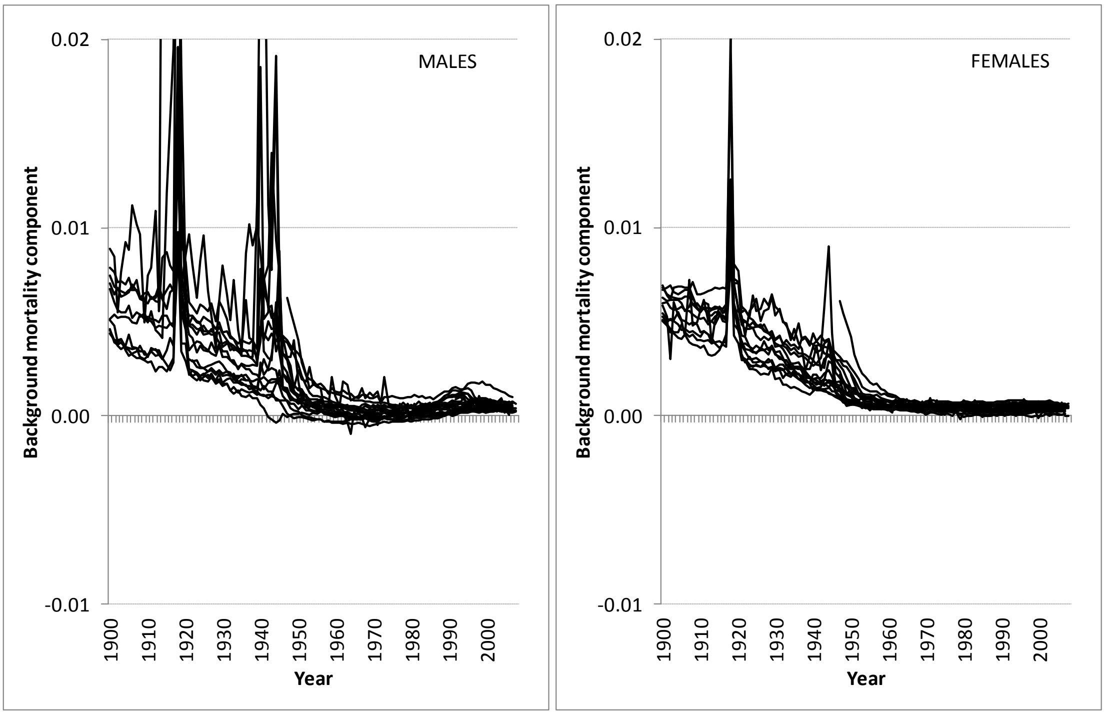
During the 1960s to 1980s the development was rather stable what was reflected in the relatively stable level of the overall mortality at that time.

Figure 73: The senescent component of mortality at age 40, PC countries, males (left), females (right), 1900–2008



Source of data: author's calculation based on Human Mortality Database (2010)

Figure 74: The background component of mortality, NPC countries, males (left), females (right), 1900–2008



Source of data: author's calculation based on Human Mortality Database (2010)

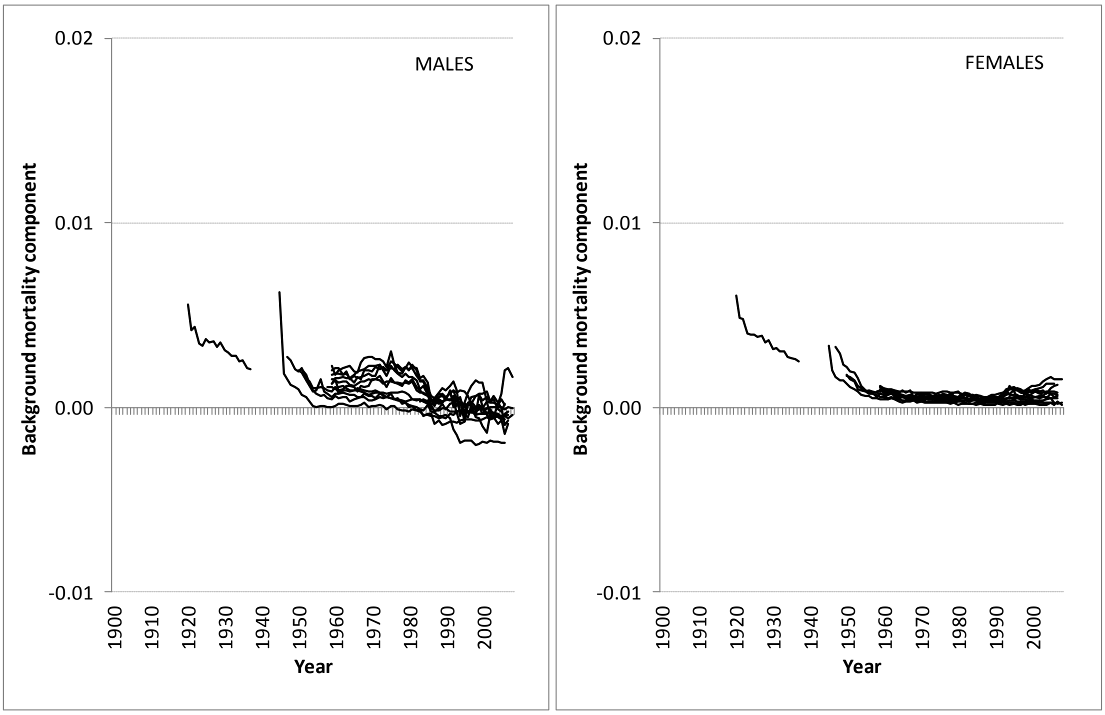
The same comparison was made for the background mortality component (Figure 74). It is the component which is independent on age, equal to the parameter *A* in the Gompertz-

Makeham model. In the NPC countries again, there the general decreasing pattern was verified for both sexes during the 1st half of the 20th century. For both sexes the values tended to zero around the middle of the 20th century. Quite an interesting situation appeared in the case of males. One can automatically suppose that the Makeham parameter cannot decrease below zero because it represents the background mortality independent on age which is usually connected with mortality caused by external factors (accidents, external causes). That is why it could be interpreted like the "unexpected deaths", that means deaths which are not connected with ageing and their appearance cannot be forecasted or expected. But the question could be posed as how to interpret the negative values of this parameter. From the mathematical point of view there is no reason for the values of this parameter to be restricted only to the non-negative values. For the Czech Republic the negative values were confirmed already by Koschin (1989). In case of negative values Koschin (2002, p. 22) suggested the possible interpretation like the "unexpected savings of life". On the example of males in the NPC countries it could be seen that the negative values of the parameter *A* appeared only temporarily during the 1950s to 1980s. Then again the parameter values returned to the positive scale. As a result of that only temporal effect, the interpretation of Koschin (2002) then could be stated as "the unexpected savings of life at a given level of mortality" – later, thanks to the positive development, the number of unexpected savings of life decrease (the parameter *A* return to positive values) again because the savings of life are already "expected" (as a result of better mortality conditions as a whole).

In the post-communist (PC) countries, there the values of the background mortality for males also decreased to negative values but not so smoothly as for males in NPC countries – in PC countries, where data are published, there the decrease after the World War II could be seen, followed by nearly stagnation from the 1960s (in post-Soviet countries the development was even increasing than stagnating; Figure 75). Among the PC countries the decrease after the World War II was the most rapid in the Czech Republic. Values of the background parameter in several PC countries during the end of the 1980s (in the Czech Republic already from the end of the 1970s) also decreased below zero as a response to relatively rapid changes which probably took more time to be reflected in the senescent mortality component. From the comparison with the NPC countries it could be expected that reaching these negative values is also only a temporary effect. This copying of the development may be useful for example in demographic forecasts which can profit from the knowledge of similarities in developmental trends. When values of the background mortality return to the positive scale, than the senescent component will be the only possible source of future mortality improvements. In that situation the background mortality will be probably relatively stable near zero level (as it is e.g. for females in NPC countries already for several decades).

For females (Figure 75) the development of the background mortality is more similar to the development in NPC countries – values were rapidly declining from the beginning of the 20th century and in some countries later reached almost stable values near zero. The decreasing trend was replaced by nearly stagnation around the middle of the 20th century in both groups of countries. Negative values were reached only in several rare cases. Again, mostly in the post-Soviet countries, we can see the worsening development represented by the increase of the background mortality at the end of the studied period.

Figure 75: The background component of mortality, PC countries, males (left), females (right), 1900–2008



Source of data: author's calculation based on Human Mortality Database (2010)

Gender differences

In the figures of the previous part of this Thesis (Figures 72–75) also the gender differences could be described briefly and also separately for the two defined groups of states.

In the NPC countries, there the decrease of the senescent mortality of females started mostly just after the end of the World War II (as was mentioned earlier, in some countries it was sooner, in some countries later), however in e.g. Switzerland or Spain the decreasing trend could be traced already almost from the beginning of the 20th century. The decrease was mainly caused by the decline of the parameter *B* because the parameter *C* was rather increasing during the 20th century, especially in its second half. Koschin (1989) proposed the interpretation of the parameter *B* as a "zero level" of natural mortality characterizing the health of the population, the quality of environment and living conditions in general. He expected this parameter to increase in the future. Although some stabilization of this parameter for females during the 2nd half of the 20th century the parameter *B* continued to decrease at the end of the century and during the first years of the 21st century. The parameter *C* was rather increasing during the second half of the 20th century. Koschin (1989) interpreted the parameter as a measure of the homogeneity of the population (the higher the value of the parameter the higher the homogeneity of the population). He expected the parameter to grow in the future what was confirmed by the empirical data.

For males the decreasing trend started on average during the 1970s – e.g. in general some 3 decades later than for females. It was also highly influenced by the values of the parameter *B* which was rather stagnating until the middle of the 20th century and then started to decrease. In the same time values of the third parameter started to increase but the decreasing trend of the parameter *B* was more significant for the overall mortality development.

In general in this and the previous part of the text it was proved that the developmental trends were in most of the NPC countries similar to the development in Sweden illustrated by Gavrilova and Gavrilov (2011) and repeated in the Figure 71.

In most of the PC countries the decrease of the senescent mortality has not started already, for males even the opposite trend could be seen. In those countries, where the decrease of the senescent mortality has already begun, the start of that trend was almost in the same time for males as for females – for females the decrease could be dated only several years earlier than for males. The development of the senescent mortality is for both sexes the most favorable in the Czech Republic and Slovenia. Again the main reason for the described trend was the development of the parameter *B*. Nevertheless that in some countries (like Russia, Latvia, Belarus, etc.) the parameter *C* is decreasing at the end of the studied period, the increase of the parameter *B* was so important that it caused the increase of the senescent mortality as a whole (especially for males). The development of the senescent mortality is clearly connected with the historical changes and rapid socio-economic development. According to the worsening of mortality levels it could be supposed that the post-revolutionary development was more problematic in the post-Soviet countries.

Considering the background mortality, development in PC and NPC countries shows some similar features – the decreasing trend heading towards the values around zero, and in both cases the temporary decrease to negative values (especially in case of males) representing the "unexpected savings of life" during rapid changes in population and in mortality which took more time to be reflected in the decrease of the senescent component of mortality.

In the case of females the trend in both the groups of countries is even more similar – the decreasing trend leading to the stabilization at very low values – the stabilization occurred a decade earlier (on average at the end of 1950s) in NPC countries than in PC countries. Again mostly in the post-Soviet countries the negative social development stands behind the increase of the background mortality – for both sexes.

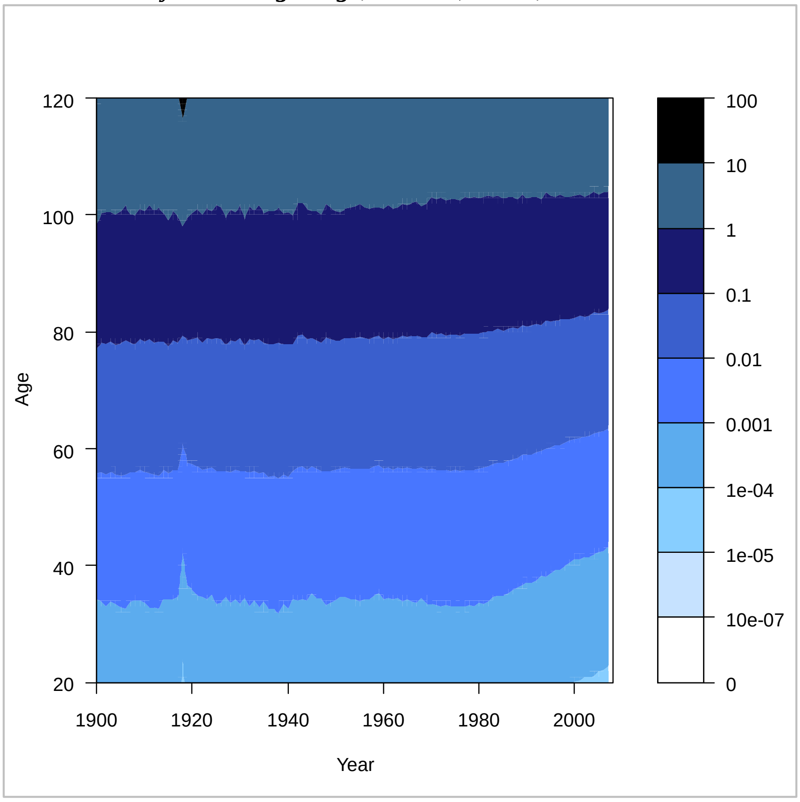
Differences of senescent mortality according to age

The research question in this section could be stated as whether there are some differences in the development of the senescent mortality according to age, i.e. whether the development of the senescent mortality was similar for all ages in a given population or whether it differs. In the previous part of this chapter the senescent mortality component was shown as calculated for the age 40, in accordance to the research of Gavrilova and Gavrilov (Gavrilova, Gavrilov, 2011). In this chapter the age-pattern of the senescent mortality is studied for selected countries.

The only way how to illustrate the development according to age and time in one picture is to use the 3-dimensional graph or a surface graph (see Chapter 2). For this chapter the second option was selected. In all those comparative graphs below, there the level of the senescent mortality is represented by various colors in a mortality surface (prepared in R-software). Results for only some selected countries are presented below because it is not possible to present here all the results. Moreover, it was confirmed that the development of the senescent mortality was similar in various ages, only small and non-significant differences could be found.

For Swedish males the senescent mortality at the age of 40 started to decline around 1980. From the graph below (Figure 76) also the development at other ages could be seen. It seems that the higher the age the slower the decrease of senescent mortality – that is the first pattern visible in the picture. The second pattern is slightly more intuitive one – it seems that the higher the age the earlier the decrease of the senescent mortality started, but the start was less sharp and braking at higher ages in comparison to lower ages. However the developmental differences do not seem to be important.

Figure 76: Senescent mortality according to age, Sweden, males, 1900–2007



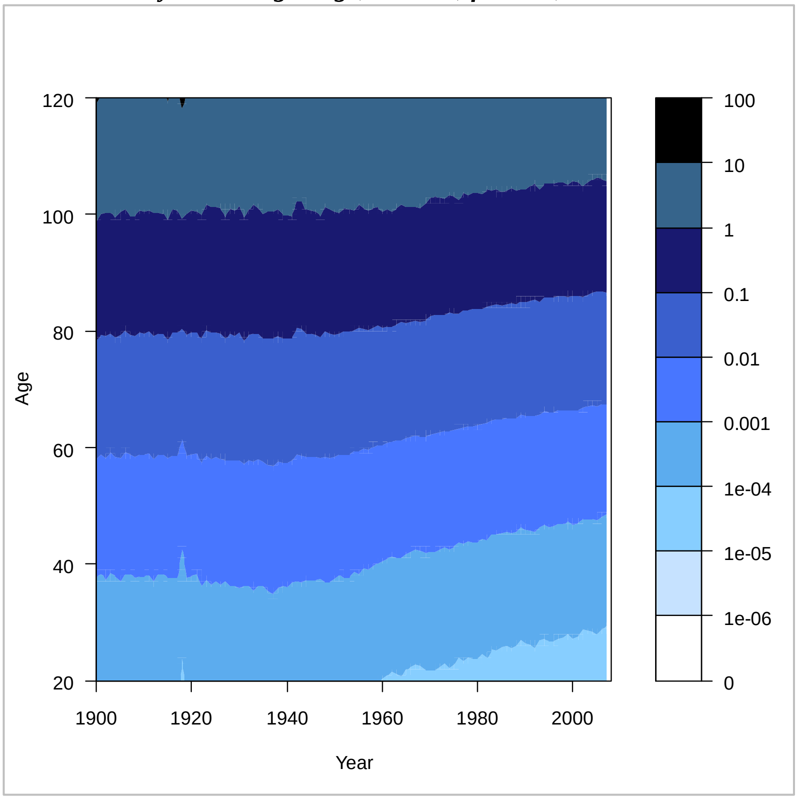
Note: Output from R software

Source of data: author's calculation based on Human Mortality Database (2010)

The same comparison was done for Swedish females. At the age of 40, the senescent mortality started to decrease in the 1950s. From the Figure 77 it could be seen that at lower ages the senescent mortality started to decrease a decade earlier. And again almost the same as for males could be verified for females too – the higher the age the smoother the break into decrease of mortality and the less rapid decline of this component of mortality.

The only exception in comparison to males is the fact that the beginning of the improvement of the senescent mortality started slightly later with increasing age. The logical conclusion is that the development (according to trend and tempo) is in Sweden for both sexes more similar with increasing age.

Figure 77: Senescent mortality according to age, Sweden, females, 1900–2007



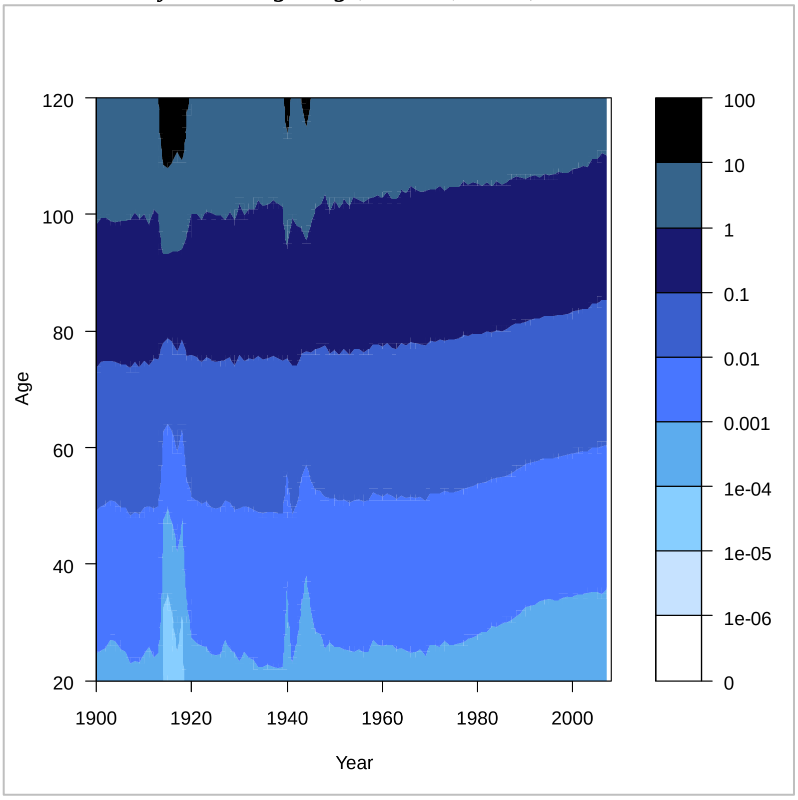
Note: Output from R software

Source of data: author's calculation based on Human Mortality Database (2010)

The described patterns could be confirmed by the example of France. For males again we can see that the lower the age the later but more rapidly the decrease of the senescent mortality started (Figure 78). At higher ages (80 and more) the senescent mortality was decreasing almost during the whole 20^{th} century (except of the World Wars). There is another interesting fact about the periods of World Wars visible – it could be noticed that the senescent mortality decreased at lower ages during the Wars and increased only at very high ages (what is rather the consequence of the applied Gomperz-Makeham method). It was the background mortality which was responsible for the rapid worsening of the mortality conditions during the World Wars. It confirms the assumption that the background mortality reflects more flexibly the rapid actual changes in the mortality conditions that do not affect the senescent mortality. It could be said that during the World Wars the importance of the "unexpected deaths" increased significantly.

Again almost the same could be stated for females (Figure 79), but for them, there is another interesting point – the stabilization of the level of senescent mortality at younger ages during the last years. After the beginning of the 21^{st} century the senescent mortality stabilized at low values at young ages (under 50) and shows no further improvement. Slightly similar pattern could be followed also for males at young ages (around 30 and less). On the other hand in the same time the senescent mortality at high ages (above 80) started to decrease even more rapidly in comparison with the previous period. The reason is the increasing value of the parameter *B* which is supposed to reflect the general living conditions and the decrease of the parameter *C* what causes the decline of the senescent mortality at higher ages.

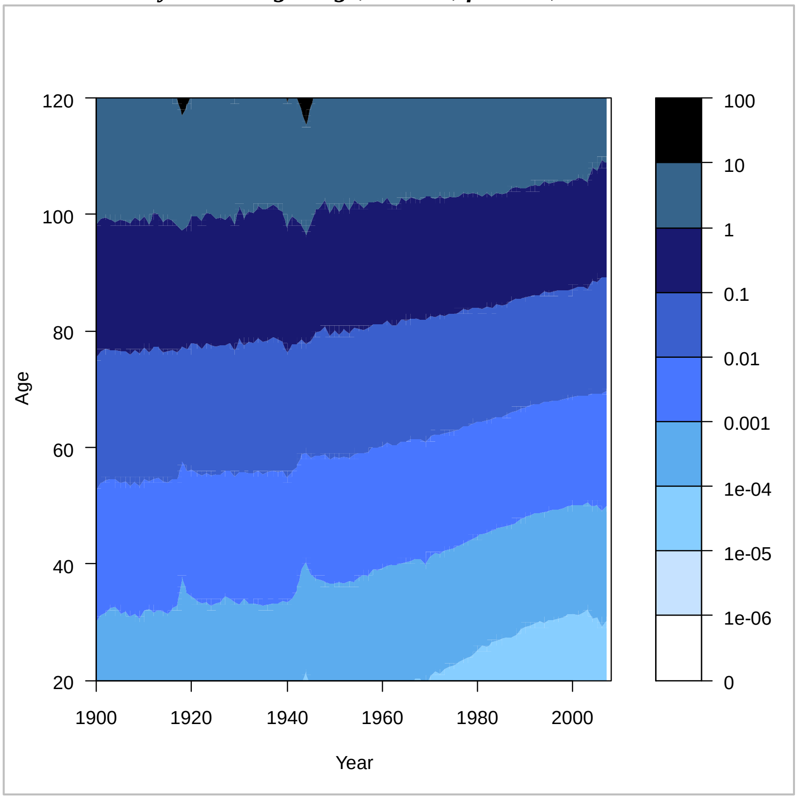
Figure 78: Senescent mortality according to age, France, males, 1900–2007



Note: Output from R software

Source of data: author's calculation based on Human Mortality Database (2010)

Figure 79: Senescent mortality according to age, France, females, 1900–2007



Note: Output from R software

Source of data: author's calculation based on Human Mortality Database (2010)

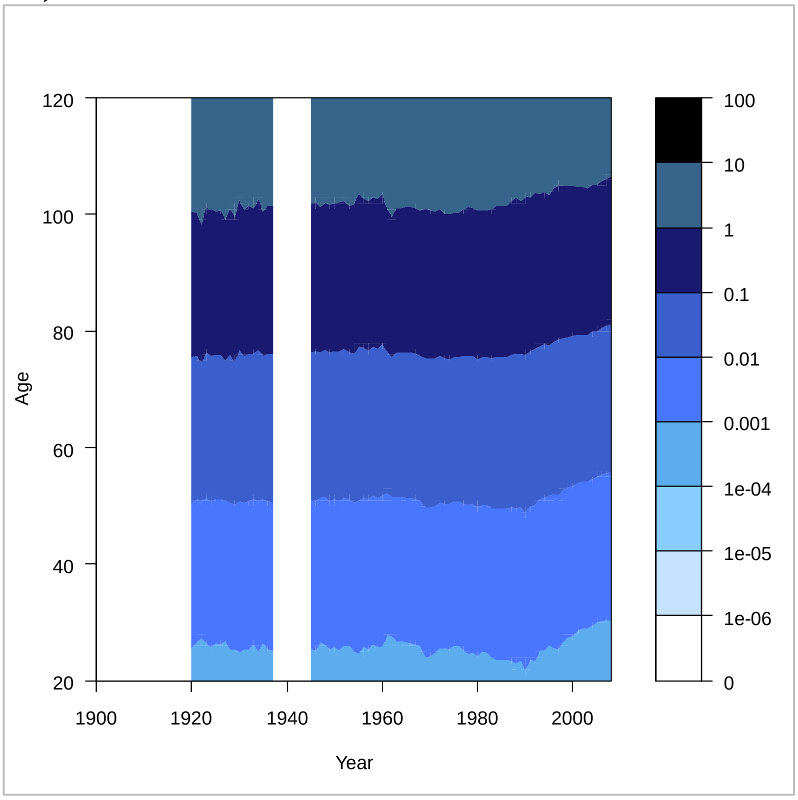
According to the proposed interpretation of the particular parameters of the Gompertz-Makeham function (Koschin, 1989) this would suggest the increasing heterogeneity of the French population during the last years and maybe also the slightly decreasing level of general living conditions.

For Swedish females it was stated that the beginning of the improvement of the senescent mortality could be dated slightly later with increasing age. For French females this was not confirmed – the higher the age was, the earlier the improvement began.

The longest time series in the group of PC countries is available for the Czech Republic (from 1920, except the years 1938–1944). For both sexes one can see that the most important changes started after the revolution in 1989 (Figures 80 and 81), but the improving tendencies were visible already during the 1980s for males (at higher ages) and almost during the entire time interval for females (except the temporary stagnation during the 1960s and 1970s which is more visible at younger ages).

The last illustration shown here is the case of Russia (Figures 82 and 83). The development of the senescent mortality shows clearly the worsening of the mortality level, especially for males. But what is more interesting, that are the differences of the development according to age. Whereas at younger ages the development seems to be rather negative or stagnating (in case of females), at higher ages, there is no mortality worsening visible, the situation seems to be more stable in comparison to younger ages. Again it is a proof of the worsening life style of adults above all at younger ages (mentioned already in the previous parts of the Thesis).

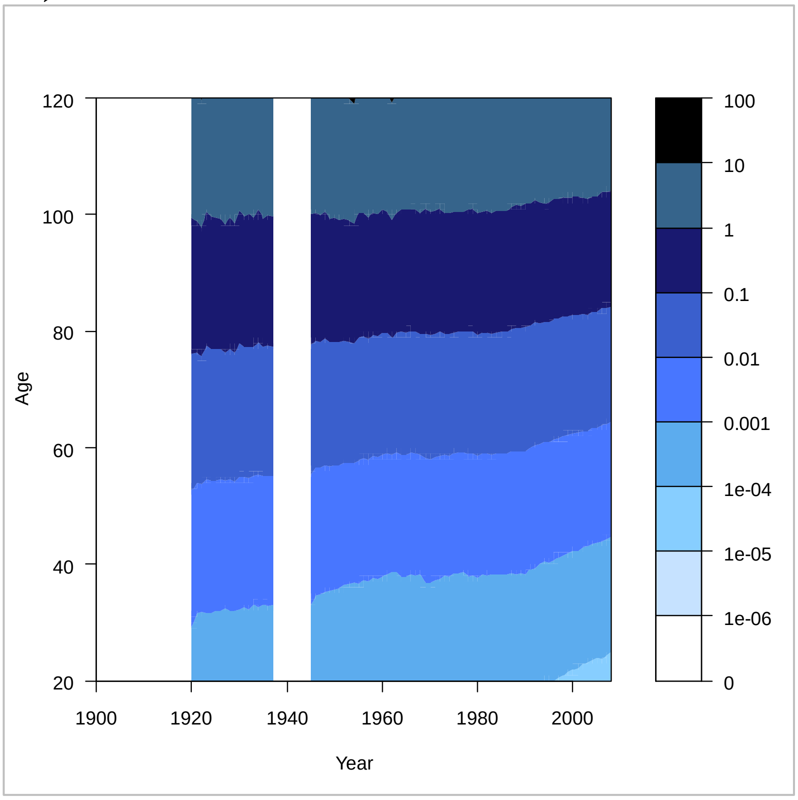
Figure 80: Senescent mortality according to age, Czech Republic, males, 1920–2008 (without the period of the World War II)



Note: Output from R software

Source of data: author's calculation based on Human Mortality Database (2010)

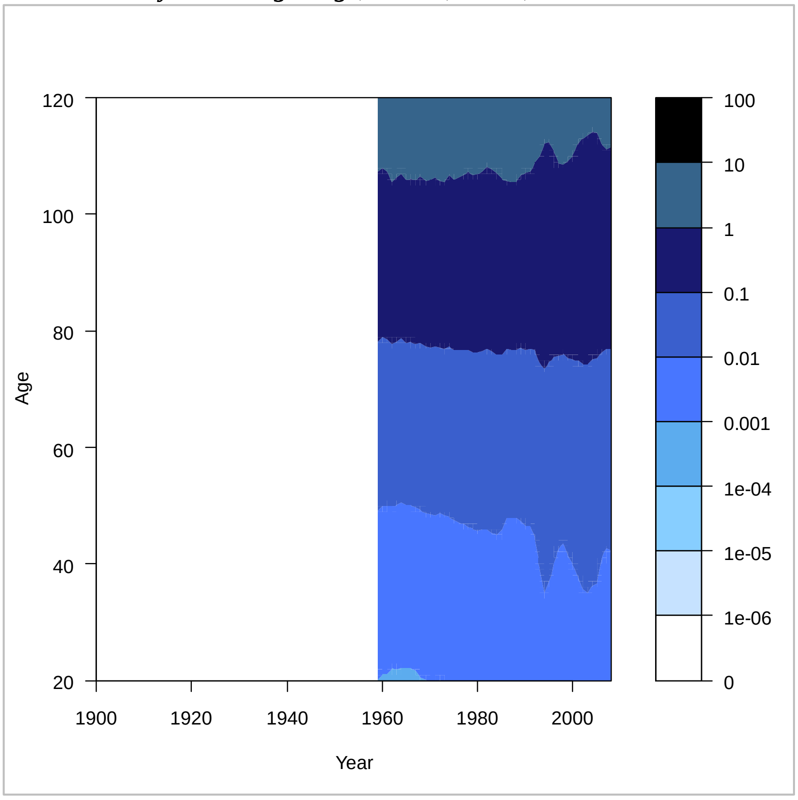
Figure 81: Senescent mortality according to age, Czech Republic, females, 1920–2008 (without the period of the World War II)



Note: Output from R software

Source of data: author's calculation based on Human Mortality Database (2010)

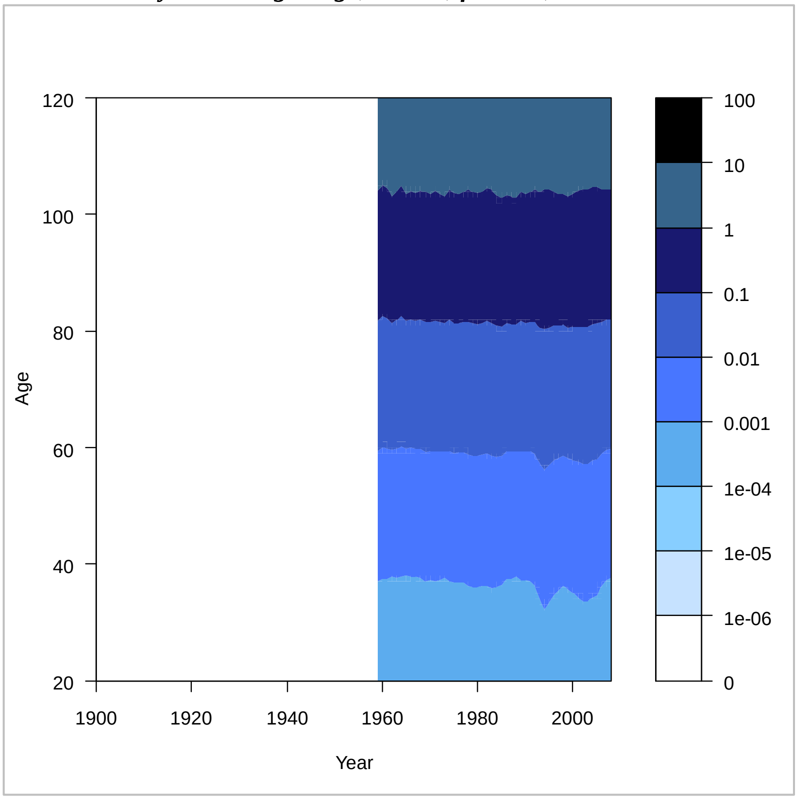
Figure 82: Senescent mortality according to age, Russia, males, 1959–2008



Note: Output from R software

Source of data: author's calculation based on Human Mortality Database (2010)

Figure 83: Senescent mortality according to age, Russia, females, 1959–2008



Note: Output from R software

Source of data: author's calculation based on Human Mortality Database (2010)

However in all the examples mentioned here it has to be kept in mind that the senescent mortality is result of the smoothing method applied to the empirical data (the Gompertz-Makeham model). That means that the observed differences could be at least partly the consequence of the applied function rather than of the real development of the data.

Model 2 – the logistic model of mortality law

Bongaarts used in his work the logistic model of mortality (Bongaarts, 2005). After the estimation of the three unknown parameters of the model he showed that the slope parameter is almost constant over time. He reached this conclusion on the basis of data from several developed countries (none of them was from the Eastern Europe). In addition to that he used the nonlinear least-square routine in STATA for the estimation of the model parameters. In our analysis the weighted nonlinear least squares method was used (as in the Model 1), so again we used a slightly different method of estimation in comparison with the original author. The weights were derived as described in the Chapter 5 and again all the calculations were done in the SAS software using the macro prepared in the fifth chapter. Based on his results, Bongaarts developed the "shifting logistic model" applicable also for forecasts (Bongaarts, 2005).

What was similar in the work of Bongaarts to the Gavrilov and Gavrilova (1979; 2011), it was the idea of dividing the adult mortality into two components – the senescent mortality (which rises with age) and the background mortality (independent on age).

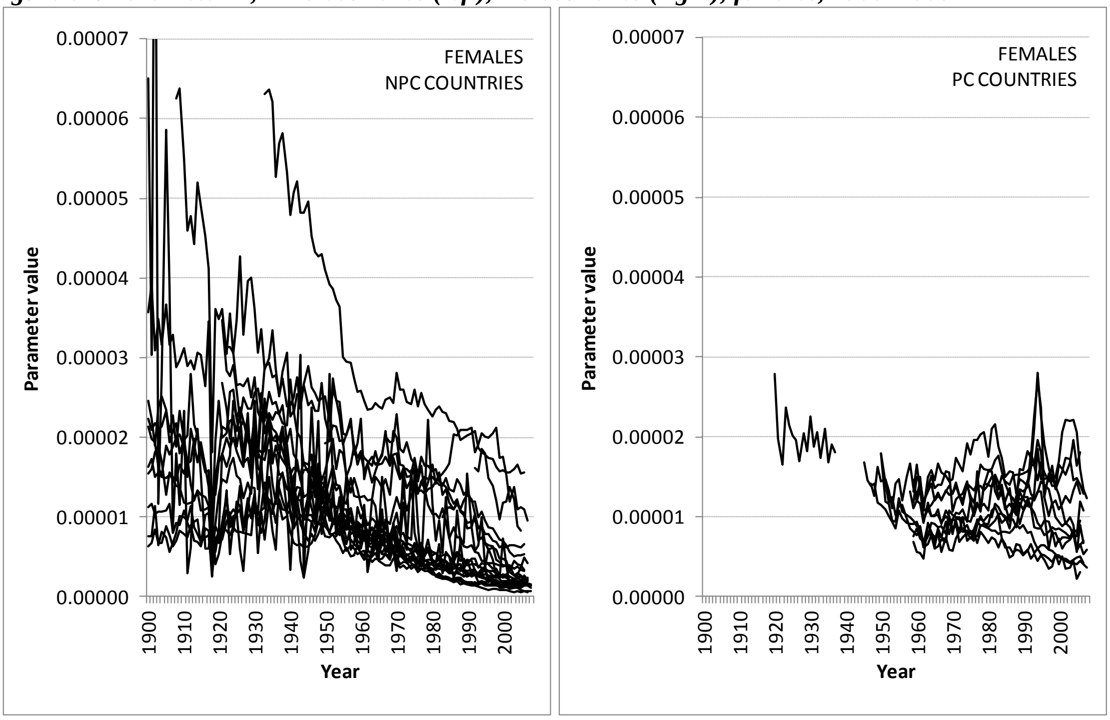
However, Bongaarts started his analysis by the study of the development of all the three parameters of the used logistic function in time (it finally showed to be more probably important than the distinction of the senescent and background mortality). He used data from the

14 developed countries during the period 1950–2000. He found relatively high variability and downward trend for the parameter *A*, constant trend and low variation for the parameter *B* and decreasing trend and a plateau at very low levels for the parameter *C* (Bongaatrs, 2005, p. 28). Based on these results he proposed his model where the parameter *B* (the slope parameter) was considered to be taken as a constant.

In this work again we would like to validate the results of the original work but also for other countries – above all the not considered in the original study (Bongaarts, 2005). Moreover, longer data time series are used where possible.

As could be seen in the Figure 84, in NPC countries the variability of the parameter *A* (the background mortality) is decreasing and also the values are lower than in the past. The picture is slightly different for females in PC countries – the trend is downward only for some of them – the Czech Republic, Slovenia or Bulgaria. In the other, there the trend is either constant or almost increasing (as it is for example in Russia).

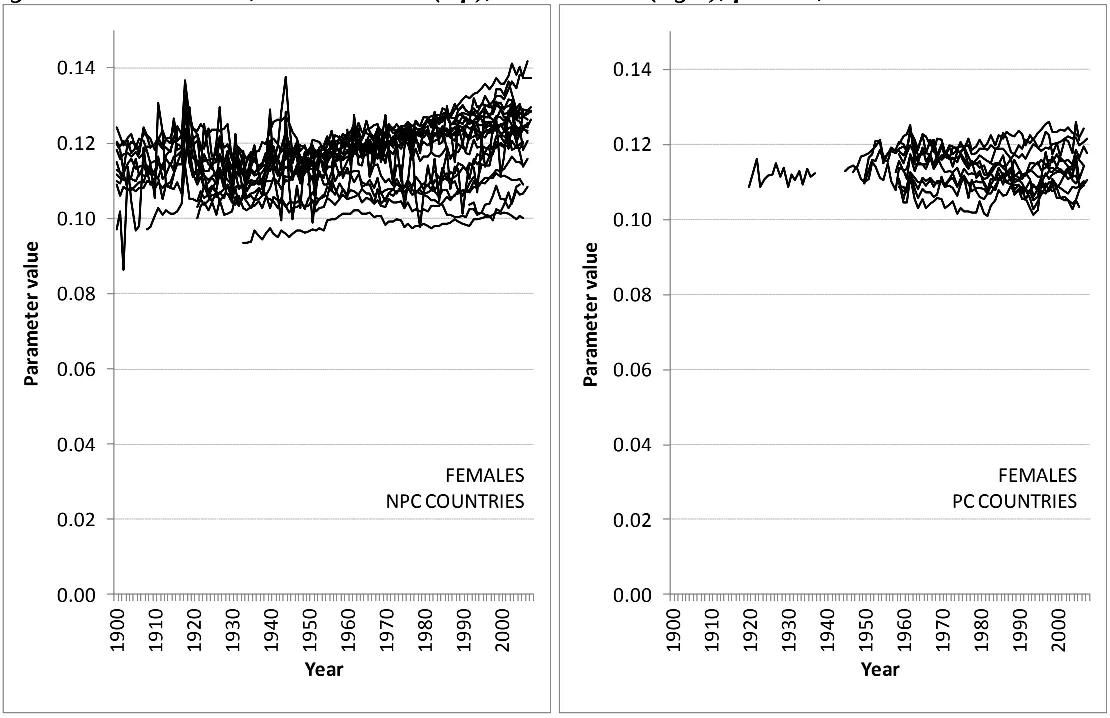
Figure 84: Parameter A, NPC countries (left), PC countries (right), females, 1900–2008



Source of data: author's calculation based on Human Mortality Database (2010)

More interesting would be the results for the parameter *B*, the slope parameter. It would show whether the most important assumption of Bongaarts (2005) was valid not only for the group of countries involved in his study. He supposed the trend to be almost constant and showed a proof in a graph for females from 1950 in the 14 developed countries. In the Figure 85 on the left side, there the longer time period and more countries (NPC countries) are showed. The trend in some of the countries could be really considered as almost constant, but in some of them it is more likely to be nearly linearly increasing (Spain, Portugal). But in most of the countries the rate of the increase is quite low. What is interesting is the result for the post-communist countries. For females it could be seen in the Figure 85 on the right side that the trend is really almost constant for all the post-communist countries in Europe. Also the values are similar to the average values of the countries without the communist history. The values of the parameter *B* in post-communist countries are only slightly lower than in the non-post-communist ones. Current values of this parameter in the Eastern Europe are almost the same as in the first group of countries some half a century ago. Therefore, it could be the proof for a very low rate of increase of this parameter and it could be concluded that in a short-range the parameter for females could be taken as constant – what is important that not only in the Western and Northern European countries, but also in the post-communist countries. When there are some exceptions, those are in countries in the Western or Southern Europe rather than in the Central or Eastern Europe.

Figure 85: Parameter B, NPC countries (left), PC countries (right), females, 1900–2008



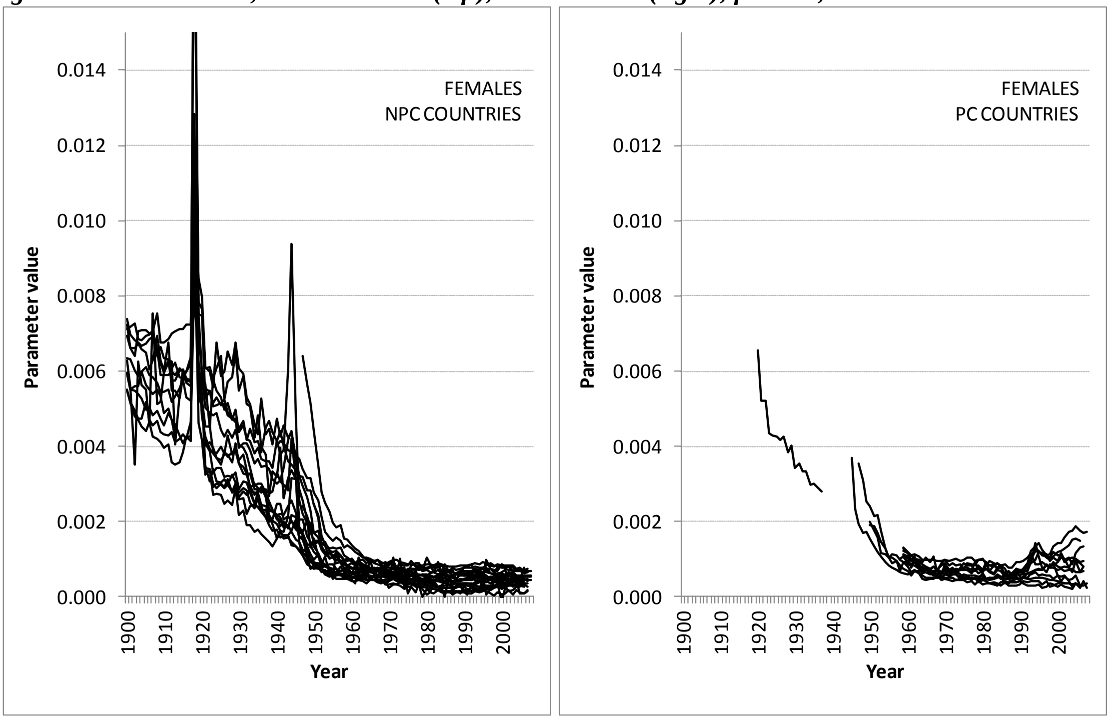
Source of data: author's calculation based on Human Mortality Database (2010)

For the third parameter, *C* (Figure 86), it was supposed that the trend was rapidly decreasing after the World War II and around the 1960s and 1970s the very low values were reached. This fact was confirmed in this Thesis too. It was also shown that the values were decreasing also before the World War II and World War I (where data were available). The parameter is also very sensitive to the actual mortality conditions (Wars or epidemics in the past).

Almost the same could be said about the development of this parameter in post-communist countries. As this parameter is very sensitive to the actual mortality conditions (it was said already), the worsening during the last two decades in some post-communist countries is represented by the growing values of the parameter *C* (in Russia or Ukraine).

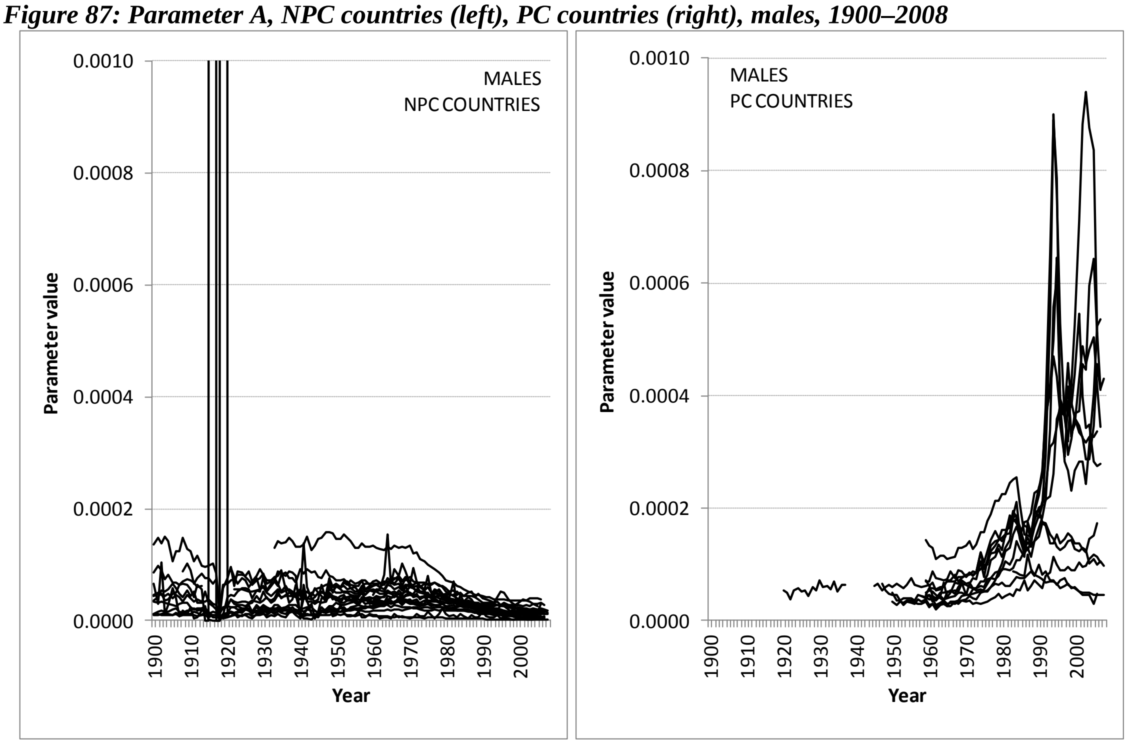
Based on the Figures 84–86 it could be said that the results of Bongaarts (2005) were confirmed for females – not only in the Western and Northern European countries (plus the most developed countries outside Europe), but also for some other non-European countries (Chile, Israel) and also the post-communist countries.

Figure 86: Parameter C, NPC countries (left), PC countries (right), females, 1900–2008



Source of data: author's calculation based on Human Mortality Database (2010)

Figure 87: Parameter A, NPC countries (left), PC countries (right), males, 1900–2008

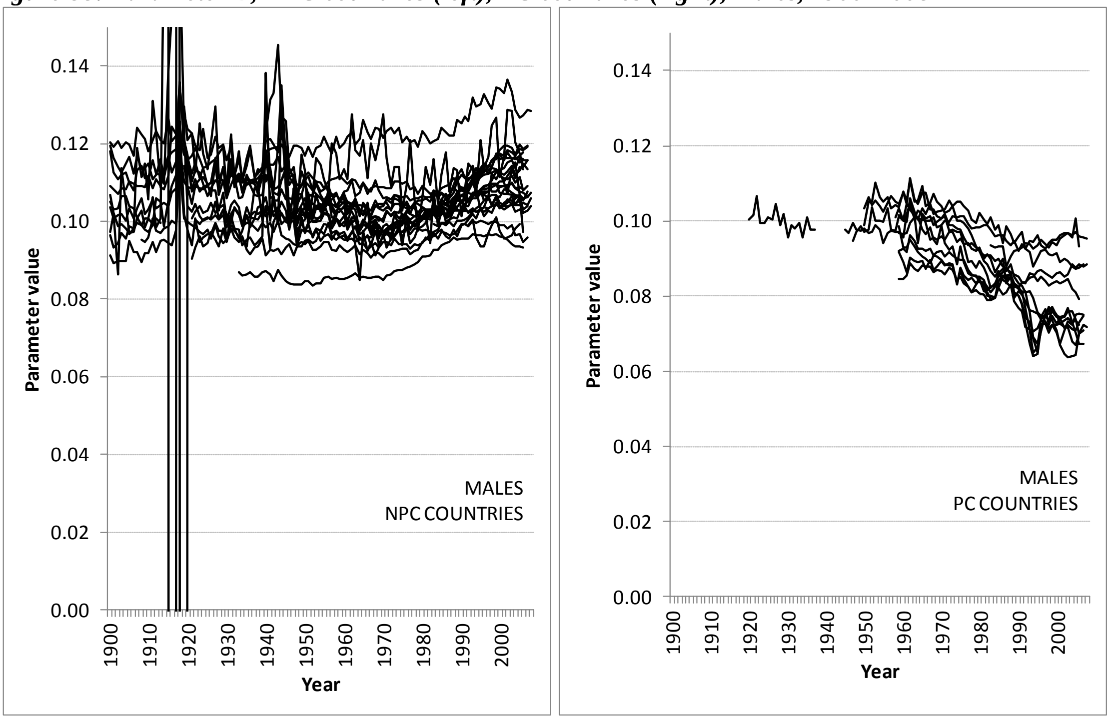


Source of data: author's calculation based on Human Mortality Database (2010)

A different result could be obtained in the case of males. The comparison with the results of Bongaarts is not possible anymore, because Bongaarts does not published in his paper the same graphs for males as he did for females (Bongaarts, 2005). From the pictures in this chapter, it could be seen that also the parameter *A* (the background mortality component) could be very sensitive (Figure 87), but only in extreme conditions (World War I for Italy and France). Of course, it has to be mentioned that also the quality of the data in that period could be the reason for the high variability of the parameter values. In the rest of the period the pattern is similar to that one of females – the decreasing values stabilizing at very low levels.

For males in the post-communist countries the differences are more visible. In some countries the worsening of mortality conditions was so huge that the values of the parameter were increasing steeply (it was the case of the post-Soviet countries).

Figure 88: Parameter B, NPC countries (left), PC countries (right), males, 1900–2008

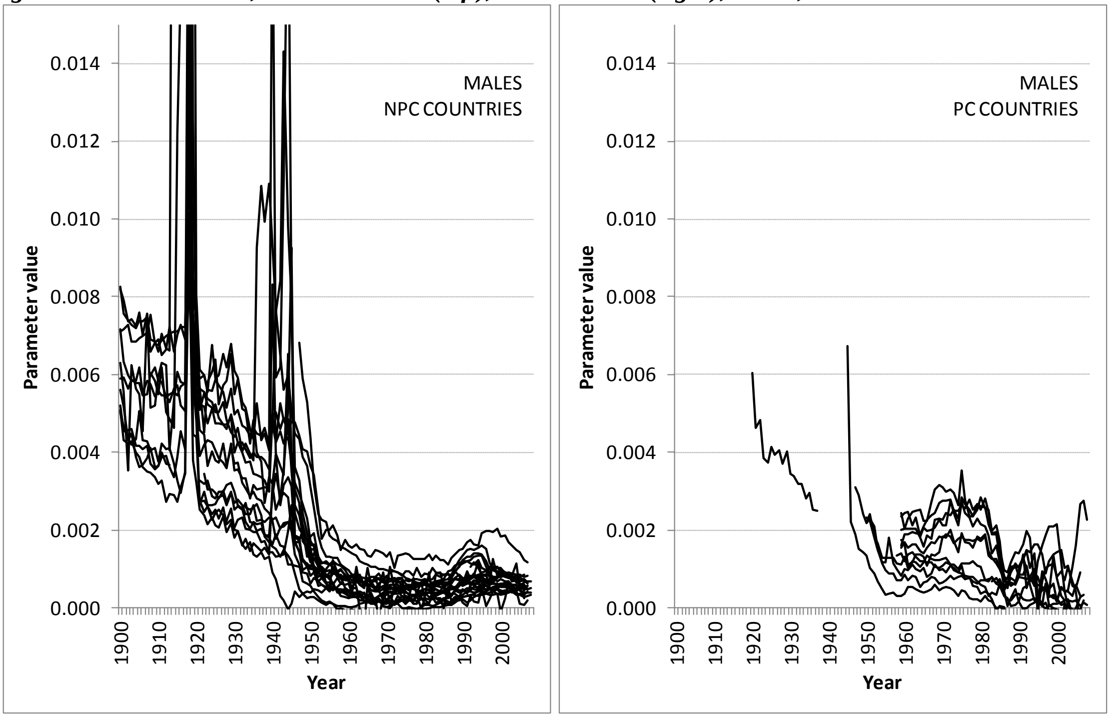


Source of data: author's calculation based on Human Mortality Database (2010)

From the development of the parameter *B* (Figure 88), it could be seen that worse mortality conditions are represented by lower values and decreasing trend of the parameter in general. Mostly the post-Soviet countries suffer from the lowest values of this parameter in the analyzed set of countries and another important fact is that the parameter is not constant in time at all. In NPC countries the trend is slightly increasing from the mid of 1970s with some stabilization at the end of the analyzed period. In the post-communist countries the trend was rather decreasing from the 1960s to 1980s in most of these countries reflecting the worse economic and social conditions. After the change of the political regime, the development was different in mostly post-Soviet countries (worsening of the conditions represented by the decrease of the parameter) and the other Central and Eastern European countries (the Czech Republic, Slovenia, Poland, Slovakia) where the trend become almost constant or even increasing (Slovenia, the Czech Republic).

Again this could be taken as the proof of the division of the post-communist countries into two different groups. Those could be found where the development was rather positive after the change of the regime (mostly the Central European countries) and those where the transformation of the society and economy brought rather a worsening of the general living conditions, economic conditions and also mortality.

Figure 89: Parameter C, NPC countries (left), PC countries (right), males, 1900–2008



Source of data: author's calculation based on Human Mortality Database (2010)

For males in the NPC countries again, there the expected trend of the parameter *C* was proved (Figure 89), as in the case of females. The stabilization at very low values started around the middle of the 20th century. Slightly higher values could be seen during the last decades in Portugal only. In the post-communist countries the trend seems to be similar, but there are almost no data for the first half of the 20th century, so the long-term trend could not be studied. Again the values of the parameter are slightly higher than in the NPC countries as a consequence of still worse mortality conditions in those countries. During the 1960s and 1970s even the worsening of the conditions is visible through the development of the parameter *C* in most of the PC countries. Almost the same worsening was repeated during the 1990s in some countries and in Russia again after 2000.

8.3.4 Conclusions of the decomposition

In the text above, there the two approaches of Gavrilov and Gavrilova (1979, 1991), Gavrilova and Gavrilov (2011) and Bongaarts (2005) were applied. The aim was to validate their results for more countries and longer time interval and also to find the general developmental trends. Also the post-Soviet and post-communist countries in Europe were involved into the analysis.

In the first model (based on the application of the Gompertz-Makeham formula) it was proved that the development of the two components of mortality (the senescent and background mortality) is similar in both groups of countries (post-communist and non-post-communist). It was shown that the background component (parameter *A*) can temporarily decrease below zero as a consequence of the improvement of mortality conditions – based on Koschin (2002) we interpreted the fact as "the unexpected savings of life in the given level of mortality". When the total mortality conditions stabilize at the generally better level, the "unexpected savings" become “expected” and the values of the background parameter returns to the positive scale again.

For the second (logistic) model it was proven that the expected development of all the parameters was validated for females in the non-post-communist countries. The development in the post-communist countries was almost the same. However, for males even in NPC countries the expected trends were not fully confirmed. The main assumption for the shifting logistic model was the stability of the parameter *B* in time (Bongaarts, 2005). This stability was not confirmed for males not only in the PC countries but even in the NPC countries. Therefore in the latter parts of the text this assumption will be hold only partially and finally it will be left and the proposal of the “age-specific shifts” will be done.

8.4 The shifting mortality hypothesis

As was stated at the beginning of this chapter, the grounds for this part are taken from the work of Bongaarts (2005). In his paper, Bongaarts used the logistic formula for estimation of the force of mortality. After the estimation of the three unknown parameters he noticed that one of them, the slope parameter *B*, was almost constant in time for the analyzed group of countries, as was described above. This result became a root of his proposal of the “shifting logistic model”. This model is applied only to the senescent component of mortality and takes into consideration only the adult ages.

In this part of the text the methodology of the estimation of the amount of the shift will be described briefly (according to Bongaarts, 2005) and the methodology will be applied to several analyzed countries. Results will be commented briefly and then the proposal of the age-specific shifts will be introduced. In that part, the assumption of the constant slope parameter will be neglected. Our own description of the shifting mortality process will be stated based upon the results of the age-specific shifts.

8.4.1 Methodology

In his analysis, Bongaarts (2005) used the logistic formula of the senescent mortality (i.e. without the background parameter *C*):

$$\mu_{x,t}^{S} = \frac{A_t * \exp(B_t * x)}{1 + A_t * \exp(B_t * x)},$$

where A_t and B_t are the parameters of the model. When we accept the condition that the slope parameter B_t is stable in time *t* the formulation simplifies to

$$\mu_{x,t}^{S} = \frac{A_t * \exp(B * x)}{1 + A_t * \exp(B * x)},$$

where *B* is a constant, A_t is the only time-variant parameter and *x* is age as usually. Bongaarts did not interpret the changes in the senescent mortality as falling and rising but rather as shifting to higher or lower ages over time. As a result “the senescent force of mortality at age *x* in year *t* is identical to the value in an earlier year t_0 at age $x - S(x)$ except around age 0” (Bongaarts, 2005, p. 29). So the formula could be rewritten as

$$\mu_{x,t}^{S} = \frac{A_{t0} * \exp\,(B * (x - S(t)))}{1 + A_{t0} * \exp\,(B * (x - S(t)))},$$

where $S(t)$ represents the amount of the shift measurable in years of age between time t_0 and t for $x > S(t)$. It was shown there that the amount of the shift could be estimated as (*ibid.*):

$$S(t) = -\frac{\ln\left(\frac{A_t}{A_{t0}}\right)}{B}$$

and the basic shifting assumption then could be described as

$$\mu_{x,t}^{S} = \mu_{x-S(t),t0}^{S} \,.$$

Bongaarts tested these formulas on a group of countries and summarized that the “shifting logistic model provides a good general description of age patterns of adult mortality in many countries for the past half century” (Bongaarts, 2005, p. 30).

8.4.2 Results and its evaluation

In this part of the chapter we continue with the analysis based on the results of the previous section where the values of the three unknown parameters were estimated by the nonlinear regression with weighted generalized least squares. At first we selected the time period in which the shift would be analyzed, that means we selected the length of the time period (Δt) and for all years t it could be written

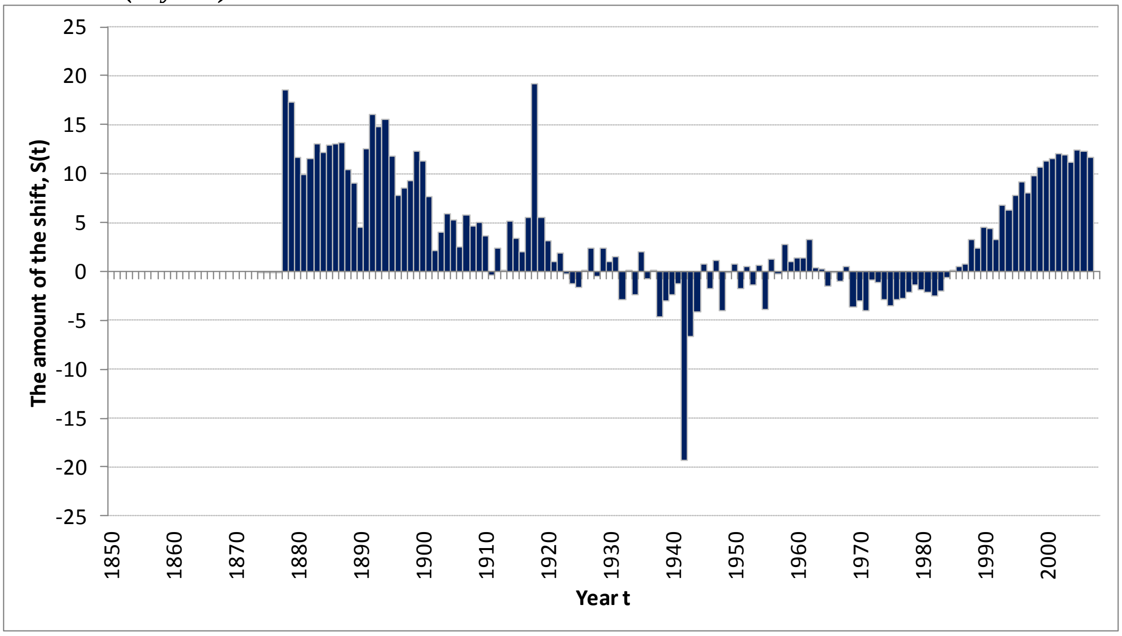
$$t_0 = t - \Delta t.$$

According to the previous equation and the equations proposed by Bongaarts we calculated the amount of the shift of the senescent mortality, $S(t)$, which occurs during the time period of the length Δt for the empirical data.

The time interval was selected as $\Delta t = 25$ as the value approximately corresponding with the length of a generation, then we calculated the amount of the shift which occurs during the 25 years proceeding to the year *t*.

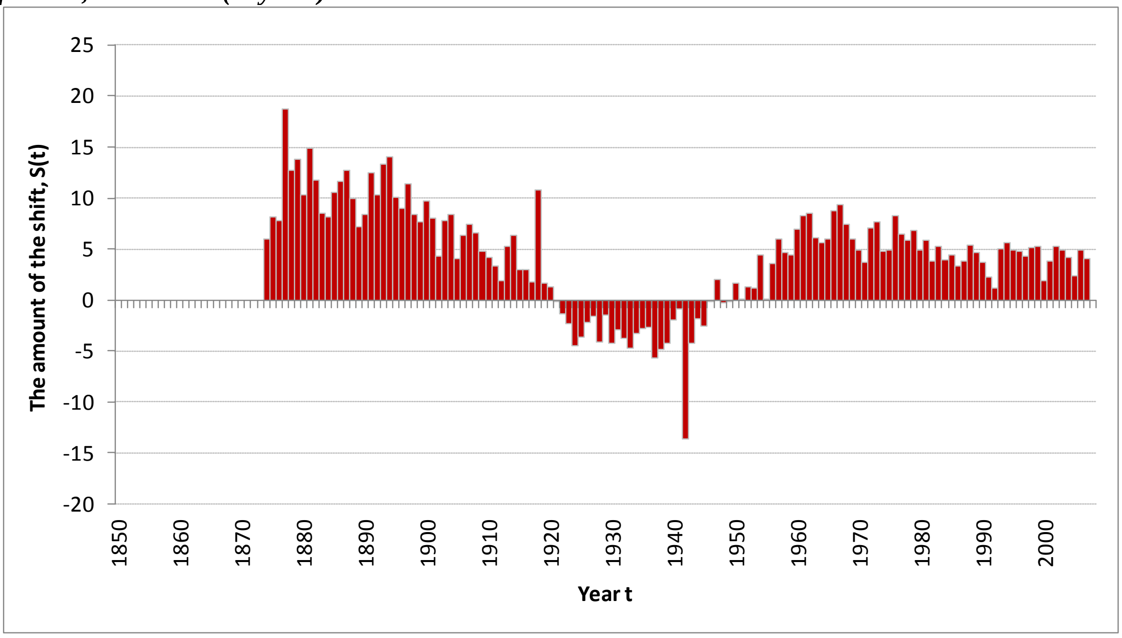
For the basic evaluation of the results we chose Sweden because of its long time-series of the mortality data. In the case of females, and also for males, it seems that there was a shifting already in the 19th century, then a temporary worsening of the mortality conditions came and the shift started again around the middle of the 20th century for females and during the 1980s for males (Figures 90, 91).

Figure 90: The amount of the shift, S(t), during the 25 years long time period (Δt = 25), Sweden, males, 1850–2008 (in years)



Source of data: author's calculation based on Human Mortality Database (2010)

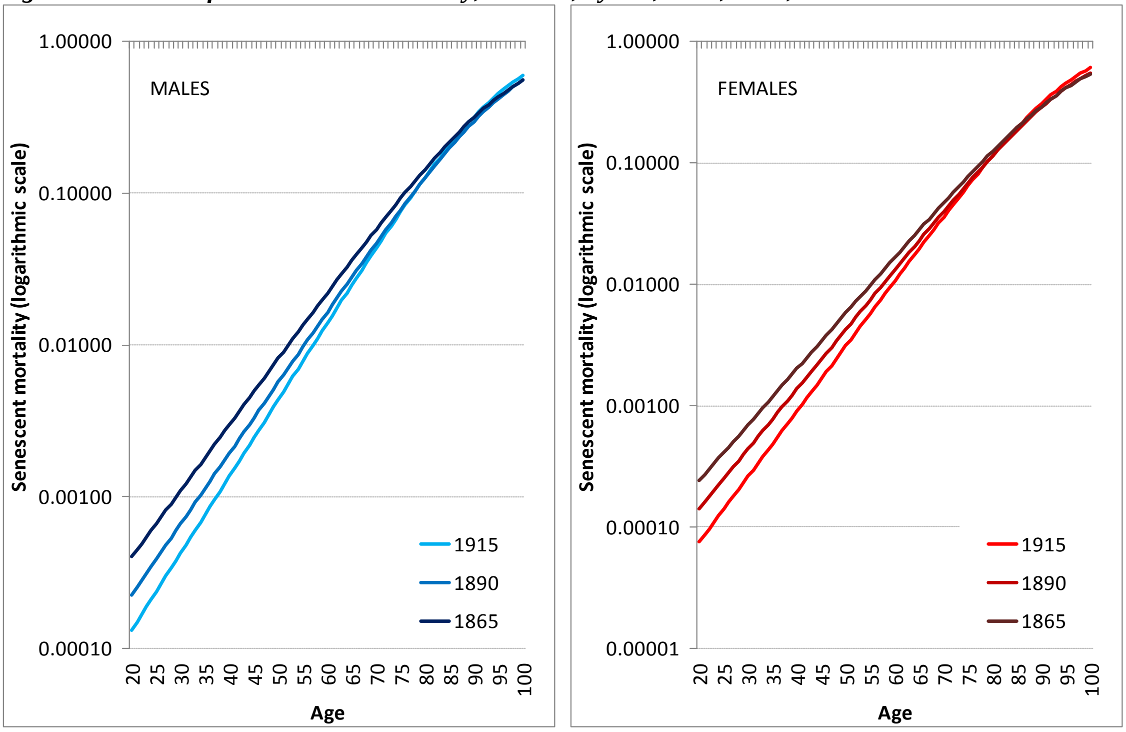
Figure 91: The amount of the shift, S(t), during the 25 years long time period (Δt = 25), Sweden, females, 1850–2008 (in years)



Source of data: author's calculation based on Human Mortality Database (2010)

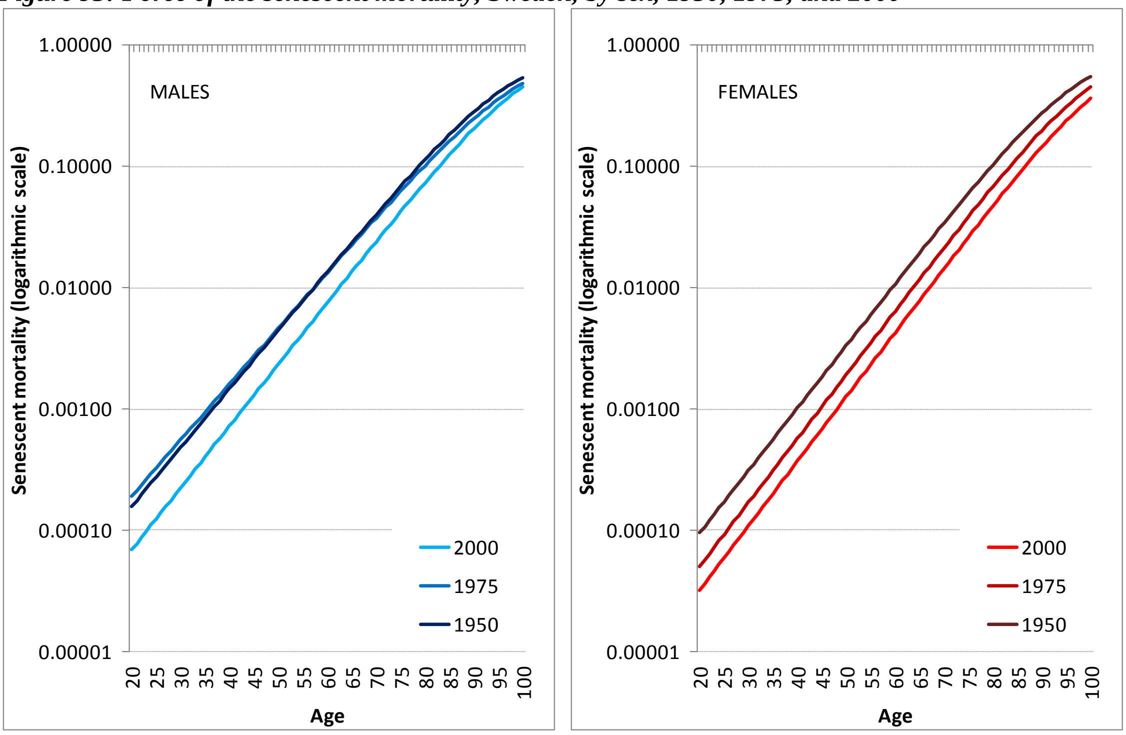
But the fact, that there was no "real shift" (in the meaning of the parallel shift constant across the age) in the 19th century, could be illustrated by the following Figure 92, where is the senescent mortality for Swedish males and females in selected ages shown (with the time lag equal to 25 years). For males the shift of the senescent mortality pattern was not confirmed even for contemporary data, for females the change of the senescent mortality could be taken as almost the parallel shift, above all in the latest data (see Figure 93). So the results in the previous figures suggesting the shift of the senescent mortality curve are only a consequence of the assumption of the constant parameter *B* which clearly was not fulfilled in the past as was proved in the Sub-chapter 8.3 in this Thesis.

Figure 92: Force of the senescent mortality, Sweden, by sex, 1865, 1890, and 1915



Source of data: author's calculation based on Human Mortality Database (2010)

Figure 93: Force of the senescent mortality, Sweden, by sex, 1950, 1975, and 2000



Source of data: author's calculation based on Human Mortality Database (2010)

8.4.3 Age-specific shifts – proposal of the estimation

Based on the results from the previous part of the text we now left the assumption of the constant slope parameter *B* behind. We will propose an alternative measure, which was called the "age-specific shifts". It will be shown that the estimation of this shifts is not much more complicated than the Bongaart's *S(t)*. Moreover, later on the presence of the mortality shifting could be verified through the values of the age-specific shifts.

When we suppose the mortality shifting to be defined (in accordance to, for example, Kannisto, 1996) as a (nearly) parallel shift of the hazard curve (or age-specific mortality rates as stated Zureick, 2010), the shifting occurs in that time when the age-specific shifts are (almost or nearly) constant for all ages. Again only the adult ages are taken into consideration.

We start with the original formula for the senescent mortality

$$\mu^S_{x,t} = \frac{A_t * \exp{(B_t * x)}}{1 + A_t * \exp{(B_t * x)}}$$

where we hold the assumption that both parameters can develop in time.

Then we assume (in accordance with Bongaarts, 2005)

$$\mu^S_{x-S(x,t),t0} = \frac{A_{t0} * \exp{(B_{t0} * (x - S(x,t)))}}{1 + A_{t0} * \exp{(B_{t0} * (x - S(x,t)))}},$$

where $S(x,t)$ is the age-specific mortality shift which occurs to a person at age *x* at time *t* during the time period $t - t_0$.

When we hold the assumption

$$\mu^S_{x,t} = \mu^S_{x-S(x,t),t0}$$

then it can be written

$$\frac{A_t * \exp{(B_t * x)}}{1 + A_t * \exp{(B_t * x)}} = \frac{A_{t0} * \exp{(B_{t0} * (x - S(x,t)))}}{1 + A_{t0} * \exp{(B_{t0} * (x - S(x,t)))}} .$$

After some basic steps we get the equation in the form

$$S(x,t) = \frac{x(B_{t0} - B_t) - \ln{(\frac{A_t}{A_{t0}})}}{B_{t0}} ,$$

where $S(x,t)$ represent the age-specific mortality shift at age *x* and time *t*.

It is clear that there is a linear relation between the shifts and age, when the formula of the age-specific shifts is written in the form

$$S(x,t) = \frac{x(B_{t0} - B_t) - \ln{(\frac{A_t}{A_{t0}})}}{B_{t0}} = x * \frac{(B_{t0} - B_t)}{B_{t0}} - \frac{\ln{(\frac{A_t}{A_{t0}})}}{B_{t0}} .$$

Therefore, the age-specific mortality shifts decrease with age when $B_{t0} < B_t$. That means that when there is increasing trend of the values of the parameter *B*, the shifts are bigger at lower ages at the expense of higher ages. When there is decreasing trend of the values of the parameter, the shifts are bigger at higher ages (that could be seen in Figure 97 on the example of the Czech males during the last decades.

If we take B_t as equal to *B*, in other words if we take the slope parameter *B* as constant in time, we will get the original Bongaarts' formula (Bongaarts, 2005):

$$S(t) = -\frac{\ln\left(\frac{A_t}{A_{t0}}\right)}{B}.$$

It is clear now, that the assumption (or better, the requirement) of constant parameter *B* can be only rarely fully fulfilled for real populations. On the top of that, the constant parameter *B* is not really the assumption for the measuring of the shift amounts, but rather it is the consequence of the real mortality shifting in the population (if it occurs). We cannot call the development as "shifting" when we only suppose the parameter to be constant, but rather we can identify the eventual shifting thanks to the constant values of the *B* parameter (estimated from the data, not only assumed). In the further analysis we suppose the slope parameter *B* to be variable in time.

Only for completeness we derived also the age-specific shifts when the Gompertz-Makeham formula is applied instead of the logistic one. Again we suppose the initial relationship

$\mu^S_{x,t} = \mu^S_{x-S(x,t),t0}$, where

$\mu^S_{x,t} = B_t * \exp\left(C_t * x\right)$ and

$\mu^S_{x-S(x,t),t0} = B_{t0} * \exp\left(C_{t0} * (x - S(x,t)\right)$.

So we have the initial assumption in a form

$B_t * \exp(C_t * x) = B_{t0} * \exp\left(C_{t0} * (x - S(x,t)\right)$ and simply we get

$$S(x,t) = \frac{-\ln\left(\frac{B_t}{B_{t0}}\right) + x(C_{t0} - C_t)}{C_{t0}} = \frac{-\ln\left(\frac{B_t}{B_{t0}}\right)}{C_{t0}} + x * \frac{(C_{t0} - C_t)}{C_{t0}}$$

what is the estimate of the age-specific shifts of mortality between the years t_0 and *t* if we use the Gompertz-Makeham formula for the fitting of the values of the hazard function. Again it is clear that there is a linear relationship between the age-specific shifts and age. And also in this case it could be summarized that when there the mortality shifting occurs in the real data it could be identified through the constant values of the parameter *C*.

8.4.4 Review of the basic results of age-specified shifts of mortality

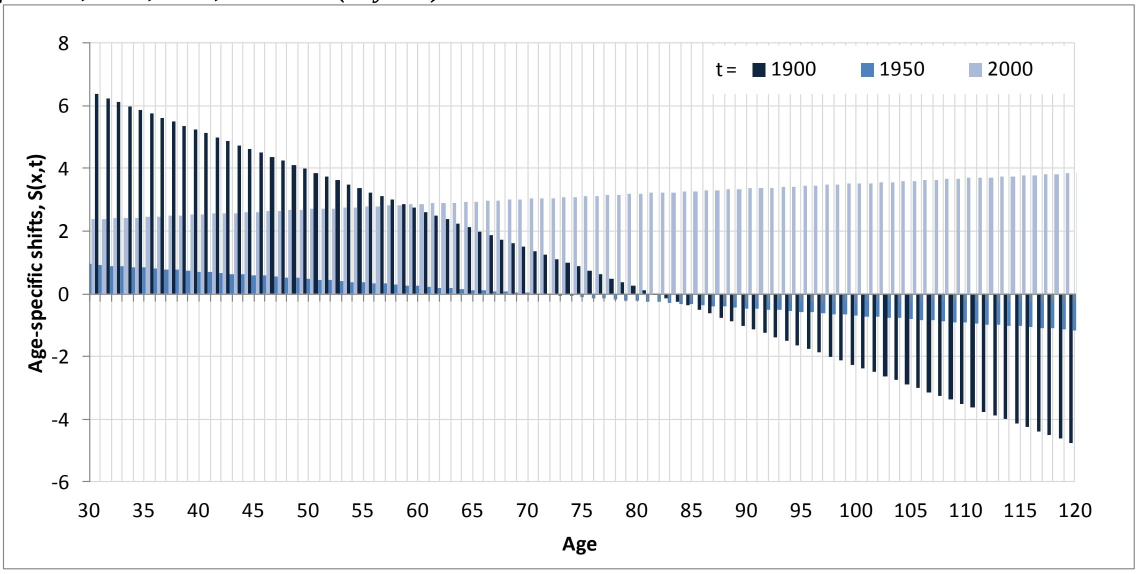
In our analysis we used only the logistic mortality model so that the results are comparable with the results of Bongaarts (2005) and his estimate of the mortality shift *S(t)*.

According to the formula derived above

$$S(x,t) = \frac{x(B_{t0} - B_t) - \ln\left(\frac{A_t}{A_{t0}}\right)}{B_{t0}}$$

we estimated the age-specific shifts for all ages above 20. For the purpose of the analysis we assumed the maximum attainable age to be equal to 120 (values of the hazard function are extrapolated by the used mortality law).

Figure 94: Age-specific mortality shifts, S(x, t), during the 25 years long time period (Δt), Sweden, females, 1900, 1950, and 2000 (in years)



Source of data: author's calculation based on Human Mortality Database (2010)

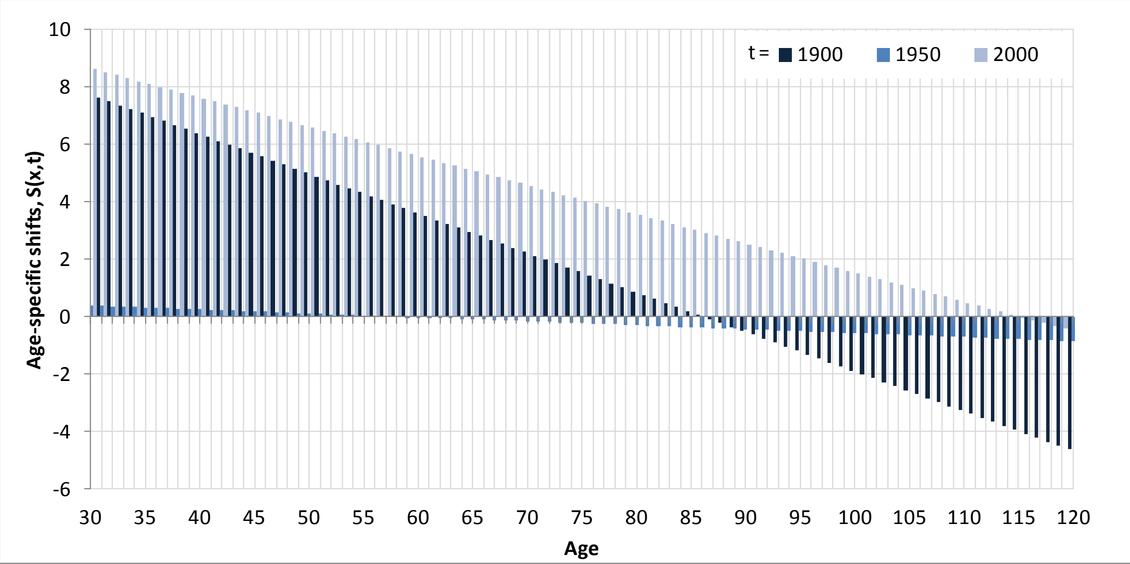
We decided to use the same time lag (the difference between time *t* and time t_0) as in the previous case, equal to 25 years and first of all we demonstrate the results for the case of Sweden (to be comparable with the previous results). Because the age-specific shifts could differ in various years we selected only some years usable for the illustration.

In the Figure 94, there the estimated age-specific shifts could be observed for Swedish females in three selected years. The time lag is 25 years again, so the estimated values of the force of mortality according to age are compared with the values of the same function but estimated for a population 25 years ago. It could be seen that when we compare the mortality in 1900 with the mortality 25 years before, then huge increases could be seen in lower ages. These values of the shifts were decreasing with age linearly so above the age around 80 the age-specific shifts are estimated to be negative. Slightly different situation could be seen for the year 1950 (again in comparison to the situation 25 years ago) – the shifts are positive at lower ages (below 70) and negative at higher ager. But the absolute values of the shifts were low – only in a few ages were the shifts higher than 1 year during the preceding 25 years.

But the situation changed completely at the end of the 20th century. It could be said that what we can see in the graph for the *t = 2000* for Swedish females, it could be almost called the mortality shifting because the age-specific shifts are almost constant with age. Of course, there is a slow increasing trend with age, so that the shifts are higher at higher ages, but if we concentrate mostly on the middle ages, the absolute increase of the shifts is not so significant.

If we took the time lag used in the analysis, 25 years, almost equal to the length of one generation, it could be summarized that Swedish females today has almost the same mortality (measured by the hazard function) as their mothers had when they were some, on average, 3 years younger.

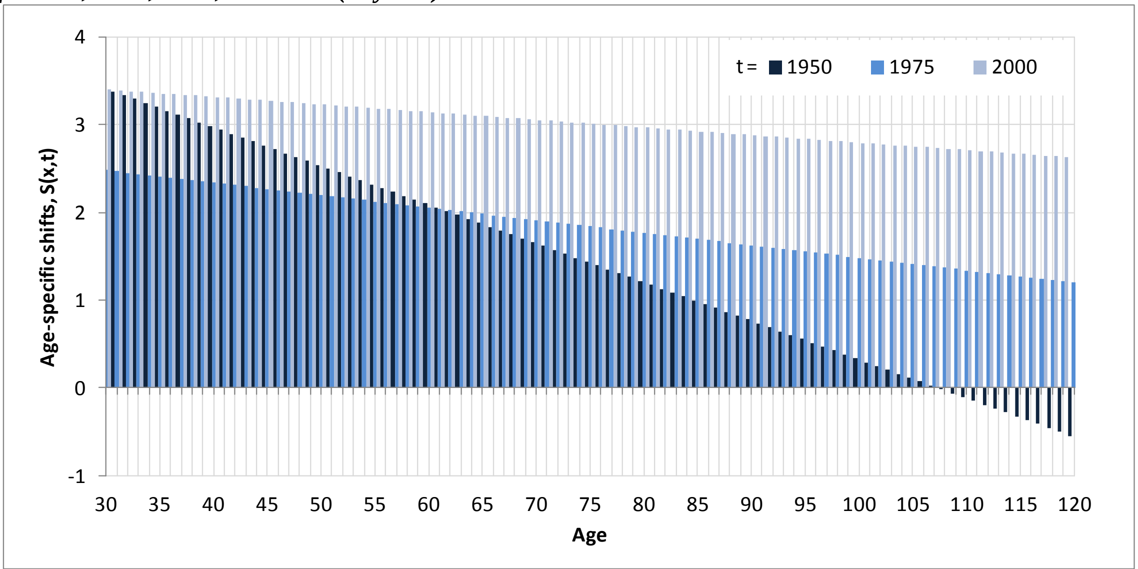
Figure 95: Age-specific mortality shifts, S(x, t), during the 25 years long time period (∆t), Sweden, males, 1900, 1950, and 2000 (in years)



Source of data: author’s calculation based on Human Mortality Database (2010)

For males, the situation is much more different. According to the Figure 95 we can suppose that Swedish males are still experiencing the improvement of the mortality conditions but the shifts are not the same at all ages – still the shifts are significantly higher at younger ages and decrease with age.

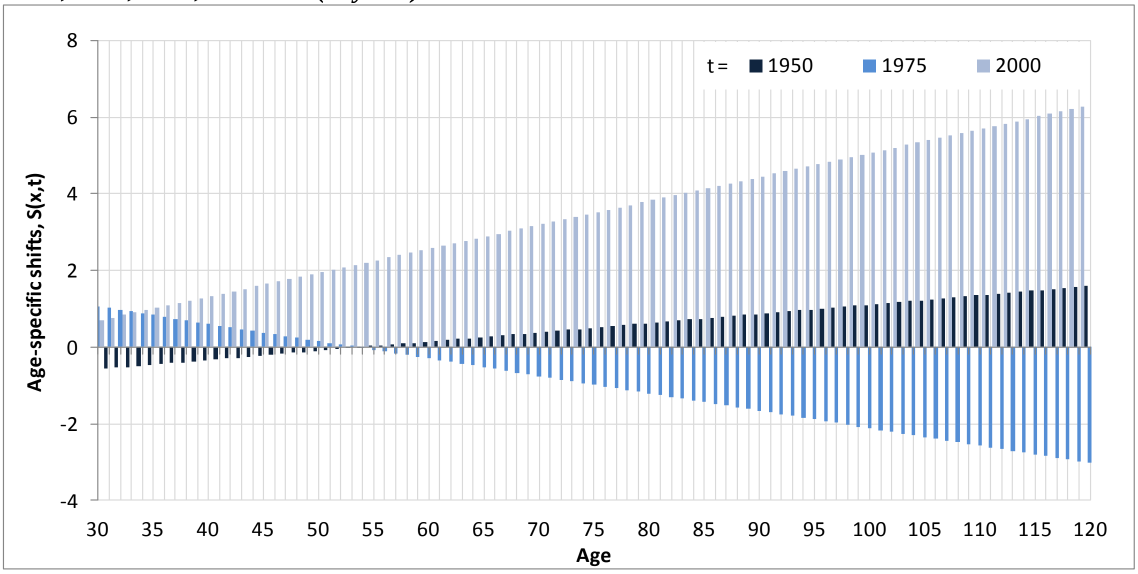
Figure 96: Age-specific mortality shifts, S(x, t), during the 25 years long time period (∆t), Czech Republic, females, 1900, 1950, and 2000 (in years)



Source of data: author’s calculation based on Human Mortality Database (2010)

The Czech Republic could be taken as a representative of the post-communist countries because of the longest time-series of reliable mortality data available. In the Figures 96 and 97, there the age-specific mortality shifts could be observed. Like in Sweden, also in the Czech Republic for females, there could be seen almost a mortality shift in the latest years. The trend of the shifts is slightly decreasing with age but on average the Czech females today experience almost the same mortality level as their mothers when they were some 3 years younger. For Czech males, there it is again almost no shift (in the meaning of the parallel shift) because although there significant mortality decrease takes place, the shifts highly differ according to age.

Figure 97: Age-specific mortality shifts, S(x, t), during the 25 years long time period (∆t), Czech Republic, males, 1900, 1950, and 2000 (in years)



Source of data: author's calculation based on Human Mortality Database (2010)

8.5 Effects of changes in mortality rates for the other table functions

There was already written enough about the shifting process in the previous parts of the Thesis. It was validated on empirical data, studied under the Bongaarts' condition or through the "age-specific shifts". In this sub-chapter we will deal with the shifting process in a theoretical and more formal way. The aim was to find out what changes in the hazard function (intensity of mortality) can theoretically lead to the parallel shift of the survival curve or of the distribution of deaths. First of all it will be studied on empirical data and then on model data representing two different laws of mortality.

8.5.1 Theoretical and methodological background

In the previous text the basic idea of the shifting mortality was expressible in the form

$\mu^S_{x,t} = \mu^S_{x-S(t),t0}$ (Bongaarts, 2005) or

$\mu^S_{x,t} = \mu^S_{x-S(x,t),t0}$ if we consider the age specific shifts proposed above.

In this way we assumed the shift as the movement of the values of the force of mortality by the age axis. In this part of the text a slightly different view will be used. We will consider the parallel declines or increases of all the values of the age-specific mortality rates and then we will study the horizontal shift of the survival curve (or of the distribution of deaths). We define two possible changes of mortality rates – a decrease or increase

- in absolute values (supposed in the "additive model" as it is used in the further text) or
- as a relative proportion from the original values ("relative model").

Both these models will be described later but first of all we have to define the basic symbols. We will study the effect of the change (absolute or relative) of age specific mortality rates between two moments of time (or generally between two populations represented by two different life tables) on other life table functions. For this purpose the mortality rate will be marked by the symbol m_x, where x denotes the completed age. Because we have to distinguish the original value of the rate (before the change) and the new value (after the change) we denote the original (not changed) value by an asterisk (m_x^*). The mortality rate after the change will be simply marked as m_x.

The life table functions will be denoted as usually, we will use the probability of survival for a person at exact age x (labeled as p_x^* for the original, unchanged table, and labeled as p_x in the new one). The relation of the change in the values of age-specific mortality rates to the survival function (marked as l_x^* and l_x) and the distribution of deaths (marked as d_x^* and d_x) will be studied in the following text.

Table 13: Basic relationships in a life table

Age	Mortality rate	Probability of surviving	Survival function	Number (density) of deaths
x_0	m_0	$p_0 = e^{-m_0}$	$l_o = 1$	$d_0 = l_0 - l_1$
x_1	m_1	$p_1 = e^{-m_1}$	$l_1 = l_o * p_o = l_o * e^{-m_0} = e^{-m_0}$	$d_1 = l_1 - l_2$
x_2	m_2	$p_2 = e^{-m_2}$	$l_2 = l_1 * p_1 = = l_o * e^{-m_0} * e^{-m_1} = = e^{-m_0} * e^{-m_1} = = e^{-(m_0+m_1)}$	$d_2 = l_2 - l_3$
…	…	…	…	…
x	m_x	$p_x = e^{-m_x}$	$l_x = e^{-\sum_0^{x-1} m_x}$	$d_x = l_x - l_{x+1}$

The basic supposed relations in the life table are illustrated by the first rows of a standard table in the Table 13 where x_0 is the initial age for the study (not necessary the age zero – we can concentrate only on the adult ages, that is why we do not handle the initial age differently from the others as is a common practice in the life table construction).

First of all we have to define the additive and relative model of the parallel change of the mortality rates.

a) The additive model

The additive model could be characterized by the decline/increase of all the age-specific mortality rates by the same absolute value, labeled Δ. That means that the new value (after the change) of the mortality rate could be written as

$$m_x = m_x^* - \Delta$$

what is the same as

$$m_x^* = m_x + \Delta.$$

That is the reason why this model was called "additional". The absolute change, Δ, is assumed to be constant across age. The effect of this parallel decline for mortality rates is illustrated in the Figure 98 on the left side.

b) The relative model

The relative model represents a situation where the decrease/increase of the original age-specific mortality rates, m_x^*, is proportional to its values. It could be written that the new value of the mortality rates (after the decline) is equal to:

$$m_x = m_x^* - k * m_x^* = m_x^* * (1 - k),$$

where k is constant across age. The absolute value of the decrease of the original mortality rates is equal to $k * m_x^*$. Because k is constant (could be also negative in case of increase of the initial mortality rates), the absolute value of the decrease/increase of the mortality rates is higher where the original mortality rates were higher. The change of mortality rates in the relative model is illustrated in the Figure 98 on the right hand side.

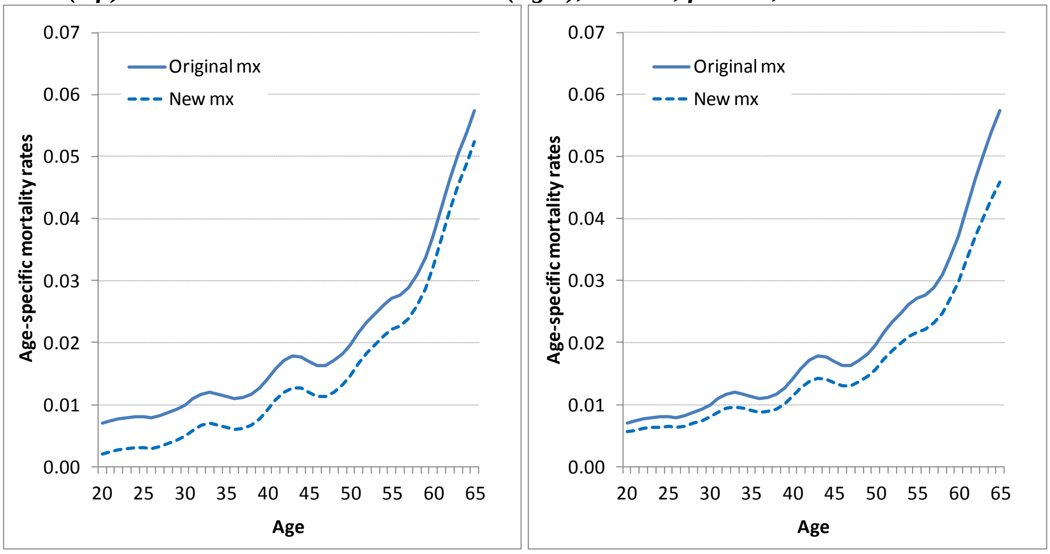
In all the illustrations below the morality rates of Swedish females in 1800 were used[15]. All the age-specific mortality rates for ages below 20 were taken as equal to zero, so as the concentration was focused only on the adult mortality. Such a historical data were selected because all the age-specific mortality rates are high enough and the considered decrease could be seen easily.

Both the changes of the mortality rates could be considered as parallel declines/increases – in the first case it is absolutely parallel (the additive model), in the second case it is relatively parallel (in the relative model). In further parts of the text, we will develop formal relationships describing the (absolute or relative) consequences of the parallel decline/increase of the age-specific mortality rates to other life table functions. In all the cases in this part we do not suppose the application of any mortality law on the data – the data are neither smoothed at any way.

---

[15] Source of the data the Human Mortality Database (http://www.mortality.org).

Figure 98: Illustration of the parallel decline of age-specific mortality rates in the additive model where* Δ *= 0.005 (left) and relative model where k = 0.20 (right), Sweden, females, 1800



Source of data: author's calculation based on Human Mortality Database (2010)

8.5.2 Impact of the parallel decline/increase of the age specific mortality rates on the survival function

In the previous part, there the two basic models, the additive and relative model, were described. According to both of them the relation of the change (decrease/increase) in age-specific mortality rates to the survival function will be derived. The question is stated as, how the original survival curve (represented by the values l_x^*) changes when there is an absolute (Δ) or relative (k) decrease (or increase) of the mortality rates.

Let's consider the additive model first and define

$$m_x = m_x^* - \Delta,$$

then it could be written for $x = 1$:

$$l_1 - l_1^* = e^{-m_0} - e^{-m_0^*} = e^{-(m_0^* - \Delta)} - e^{-m_0^*} = e^{-m_0^*} * e^{\Delta} - e^{-m_0^*} =$$

$$= e^{-m_0^*} * (e^{\Delta} - 1) = l_1^* * (e^{\Delta} - 1)\ .$$

In general for all x (where $\Delta < m_x^*$; for x where $\Delta \geq m_x^*$ the decreased mortality rates m_x were taken as being equal to zero. Theoretically, the formula could be also taken as valid for all x if we accept the possibility that m_x could be negative – as in special cases of the increment-decrement tables):

$$l_x - l_x^* = e^{-\sum_0^{x-1} m_x} - e^{-\sum_0^{x-1} m_x^*} = e^{-\sum_0^{x-1} (m_x^* - \Delta)} - e^{-\sum_0^{x-1} m_x^*} =$$

$$= e^{x*\Delta - \sum_0^{x-1} m_x^*} - e^{-\sum_0^{x-1} m_x^*} = e^{-\sum_0^{x-1} m_x^*} * (e^{x*\Delta} - 1) = l_x^* * (e^{x*\Delta} - 1)$$

Now it has been proved that the absolute change of the survival curve ($l_x - l_x^*$; Figure 99, on the left) depends on the original values of the survival curve (l_x^*), or on the original distribution of the age-specific mortality rates (m_x^*), and on the term representing the relative change ($e^{x*\Delta} - 1$) which is determined by the absolute decrease/increase of the mortality rates (Δ) and by age (x).

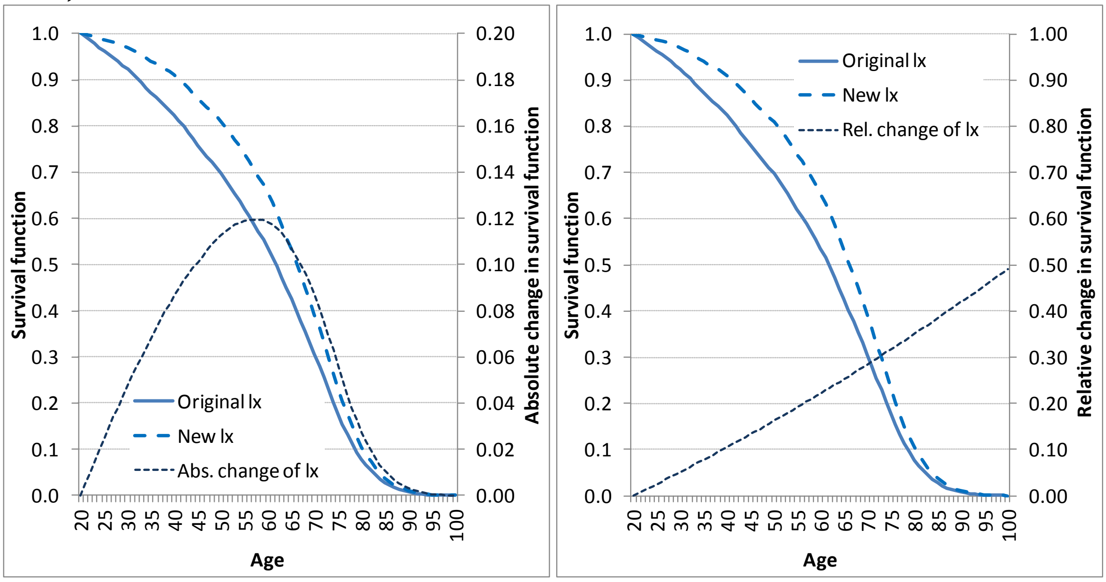
The term representing the relative change of the survival curve could be denoted as ϕ_x and so we can write

$$\phi_x = \frac{l_x - l_x^*}{l_x^*} = (e^{x*\Delta} - 1) \text{ and}$$

$$l_x - l_x^* = e^{-\sum_0^{x-1} m_x^*} * \phi_x = l_x^* * \phi_x \; .$$

The relative change of the survival curve (ϕ_x) in an additive model grows exponentially with age (Figure 99, on the right). Its rate of increase is higher for higher values of the absolute decrease of the mortality rates (Δ). What is important here is the fact that the relative change of the survival curve (ϕ_x) does not depend on the initial distribution and also on the values of the age-specific mortality rates. So when there is absolute parallel change of the initial mortality rates in the table, there would be also the exponentially (with age) increasing change of the survival curve.

Figure 99: Absolute and relative (φ_x) change of the survival curve, additive model, Δ = 0.005, Sweden, females, 1800



Source of data: author's calculation based on Human Mortality Database (2010)

We can conclude that the relative change of the survival curve (φ_x) increases with age while there is absolutely parallel decline of the mortality rates.

When we consider the relative model we can write

$$m_x = m_x^* * (1 - k)$$

and then it holds that

$$l_1 = e^{-m_0} = e^{-(m_0^* - \mathrm{k} * m_0^*)} = e^{-(m_0^* * (1-k))} = \left(e^{-m_0^*}\right)^{(1-k)} = (l_1^*)^{(1-k)}$$

and in general

$$l_x = \left(e^{-\sum_0^{x-1} m_x^*}\right)^{1-k} = (l_x^*)^{(1-k)} .$$

Then the absolute change of the survival curve could be defined as

$$l_x - l_x^* = (l_x^*)^{(1-k)} - l_x^* = l_x^* * \left[(l_x^*)^{(-k)} - 1\right] = e^{-\sum_0^{x-1} m_x^*} * \left(e^{\sum_0^{x-1} k * m_x^*} - 1\right) .$$

In case of the relative model, the absolute change of the survival curve (Figure 100, on the left) depends on the value of the relative decrease/increase of the initial mortality rates (*k*) and on the initial survival curve (l_x^*) or initial mortality rates distribution (m_x^*).

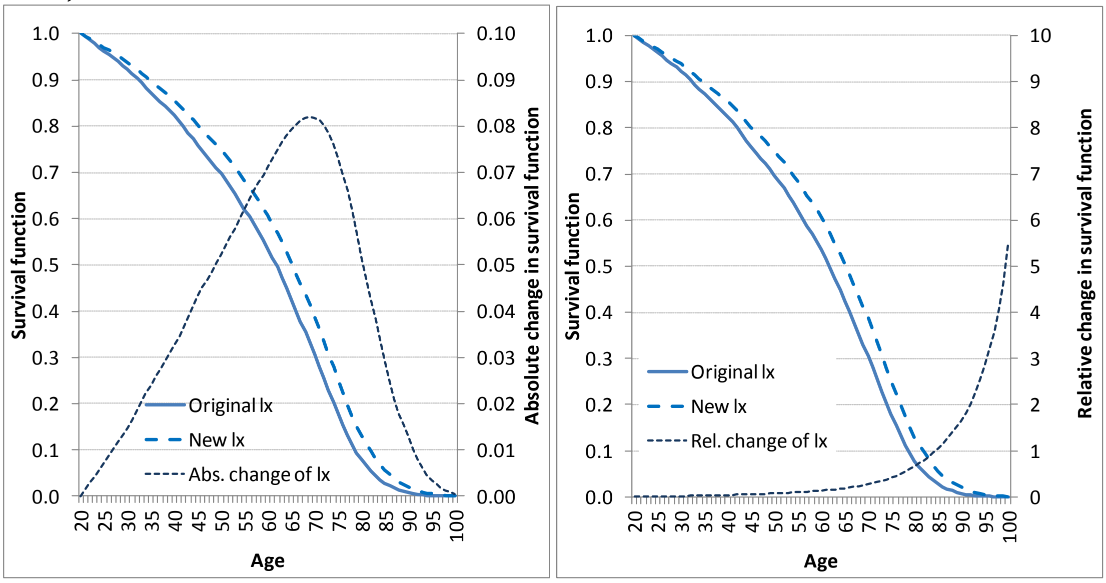
In accordance with the additional model we can easily define the relative change of the survival curve denoted by F_x:

$$F_x = \frac{l_x - l_x^*}{l_x^*} = \frac{(l_x^*)^{(1-k)} - l_x^*}{l_x^*} = e^{\sum_0^{x-1} k * m_x^*} - 1$$

As could be seen in the graphs the relative change of the survival curve increases rapidly with age at higher ages. For the group of the oldest-old in this model example, it holds true that the original curve (l_x^*) at the ages above ca 85 years would be more than doubled after the theoretical 20 per cent decrease in all the age-specific mortality rates (Figure 100 on the right).

As a consequence of the results (obtained for both the models above), the relative model represented by the proportional decrease of the mortality rates is closer to the shifting mortality hypothesis if it is understand as the horizontal shift of the survival curve (Feeney, 2006) or if it is measured by the indicators of rectangularization of the survival curve (Canudas-Romo, 2008, Zureick, 2010). It will be proved in general in the following text. In both the models the survival curve became visibly more rectangular after the decrease of the mortality rates but in the relative model the numbers of survivors increased more importantly at the highest ages. It is a natural and expected consequence of the fact that in the relative model the absolute decrease of the age-specific mortality rates was bigger at higher ages because in those ages also the initial mortality rates were higher.

Figure 100: Absolute and relative (F_x) change of the survival curve, relative model, k = 0.20, Sweden, females, 1800



Source of data: author's calculation based on Human Mortality Database (2010)

8.5.3 Impact of the parallel decline/increase of age specific mortality rates on the function of the distribution of deaths

Because the process of mortality shifting could be studied also through the changes of the probability density function, in this section the similar relationships as in the previous one will be derived for the function of the distribution (density) of deaths in the life table, denoted traditionally as d_x.

The additive model will be taken first again. If we consider

$$d_x^* = l_x^* - l_{x+1}^* = e^{-\sum_0^{x-1} m_x^*} - e^{-\sum_0^{x} m_x^*}$$

and

$$d_x = l_x - l_{x+1} = l_x^* * e^{x*\Delta} - l_{x+1}^* * e^{(x+1)*\Delta} = e^{-\sum_0^{x-1} m_x^*} * e^{x*\Delta} - e^{-\sum_0^{x} m_x^*} * e^{(x+1)*\Delta}$$

$$= e^{-\sum_0^{x-1} m_x^*} * (\phi_x + 1) - e^{-\sum_0^{x} m_x^*}(\phi_{x+1} + 1)$$

then it could be derived that

$$d_x - d_x^* = e^{-\sum_0^{x-1} m_x^*} * \phi_x - e^{-\sum_0^{x} m_x^*} * \phi_{x+1},$$

where

$$\phi_x = (e^{x*\Delta} - 1)\ .$$

The last thing in this additive model is to derive the relation for the relative change of the numbers of death, ρ_x:

$$\rho_x = \frac{d_x - d_x^*}{d_x^*} = \frac{\phi_x - e^{-m_x^*} * \phi_{x+1}}{1 - e^{-m_x^*}} .$$

In our model example the variability of ages at death clearly decreases because deaths are more concentrated around the mode (which does not change significantly) and the number of deaths at the modal age increased after the change in initial mortality rates (Figure 101).

In the relative model the situation is a bit more complicated. The difference between the initial d_x^* and values after the change of the mortality rates (d_x) could be estimated as

$$d_x - d_x^* = e^{-\sum_0^{x-1} m_x^*} * \left(e^{\sum_0^{x-1} k*m_x^*} - 1\right) - e^{-\sum_0^{x} m_x^*} * \left(e^{\sum_0^{x} k*m_x^*} - 1\right),$$

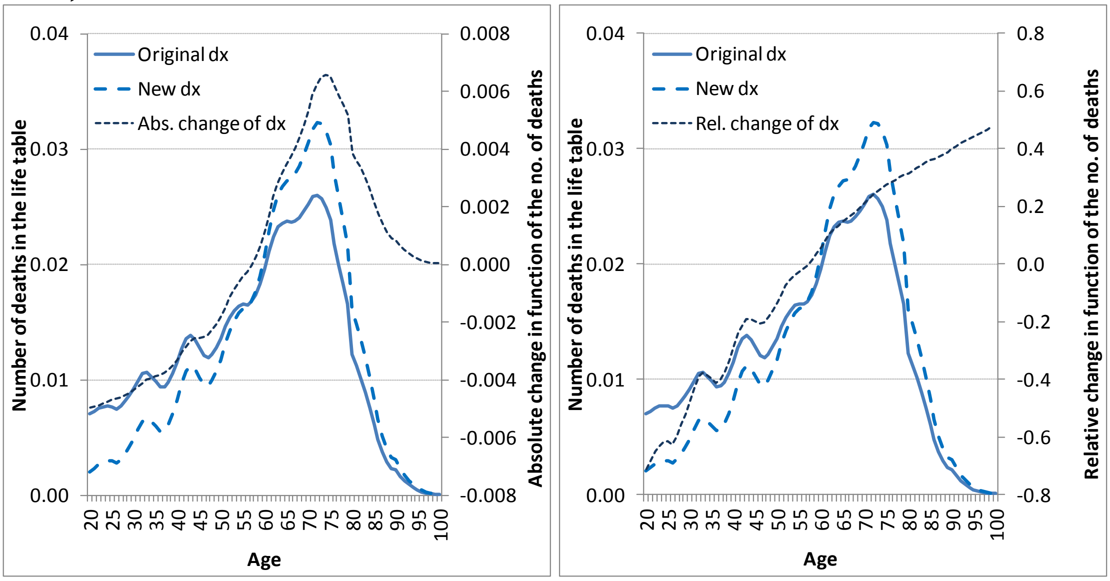
for simplification it is possible to define $k * m_x^* = \delta_x$ what is the value of the absolute change of the initial mortality rates in the relative model.

Then we have

$$d_x - d_x^* = e^{-\sum_0^{x-1} m_x^*} * \left(e^{\sum_0^{x-1} \delta_x} - 1\right) - e^{-\sum_0^{x} m_x^*} * \left(e^{\sum_0^{x} \delta_x} - 1\right) =$$

$$= l_x^* * \left(e^{\sum_0^{x-1} \delta_x} - 1\right) - l_{x+1}^* * \left(e^{\sum_0^{x} \delta_x} - 1\right) .$$

Figure 101: Absolute and relative (ρ_x) change of the survival curve, additive model, Δ = 0.005, Sweden, females, 1800



Source of data: author's calculation based on Human Mortality Database (2010)

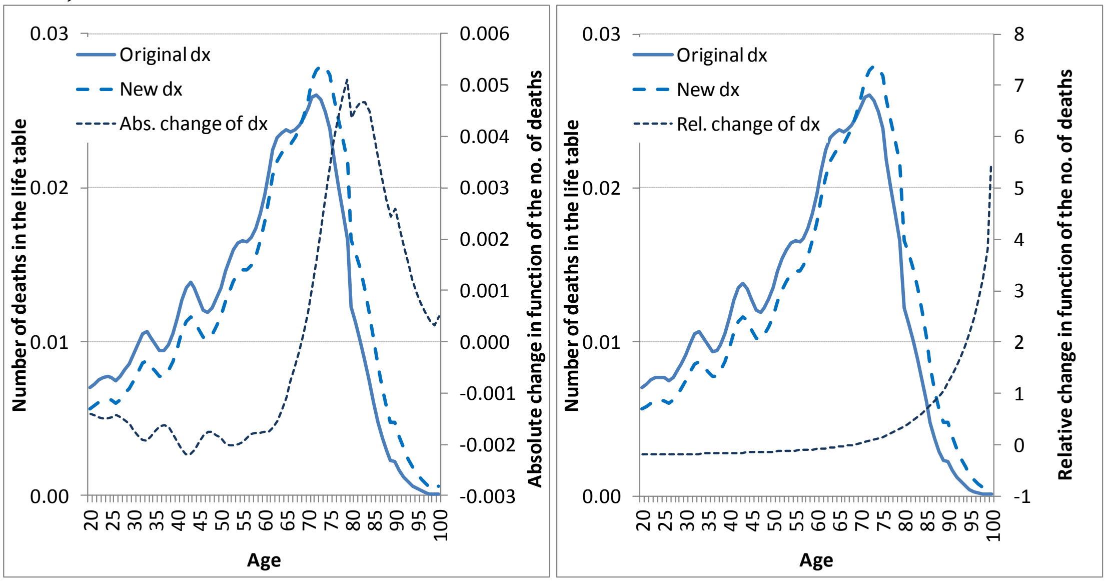
The relative difference between the initial d_x^* and values after the change of the mortality rates (d_x) could be for the relative model denoted by the letter R_x and derived as

$$R_x = \frac{d_x - d_x^*}{d_x^*} = \frac{e^{-\sum_0^{x-1} m_x^*} * \left(e^{\sum_0^{x-1} k*m_x^*} - 1\right) - e^{-\sum_0^{x} m_x^*} * \left(e^{\sum_0^{x} k*m_x^*} - 1\right)}{e^{-\sum_0^{x-1} m_x^*} - e^{-\sum_0^{x} m_x^*}} =$$

$$= \frac{l_x^* * \left(e^{\sum_0^{x-1} \delta_x} - 1\right) - l_{x+1}^* * \left(e^{\sum_0^{x} \delta_x} - 1\right)}{l_x^* - l_{x+1}^*} = \frac{\left(e^{\sum_0^{x-1} \delta_x} - 1\right) - e^{-m_x^*} * \left(e^{\sum_0^{x} \delta_x} - 1\right)}{1 - e^{-m_x^*}}.$$

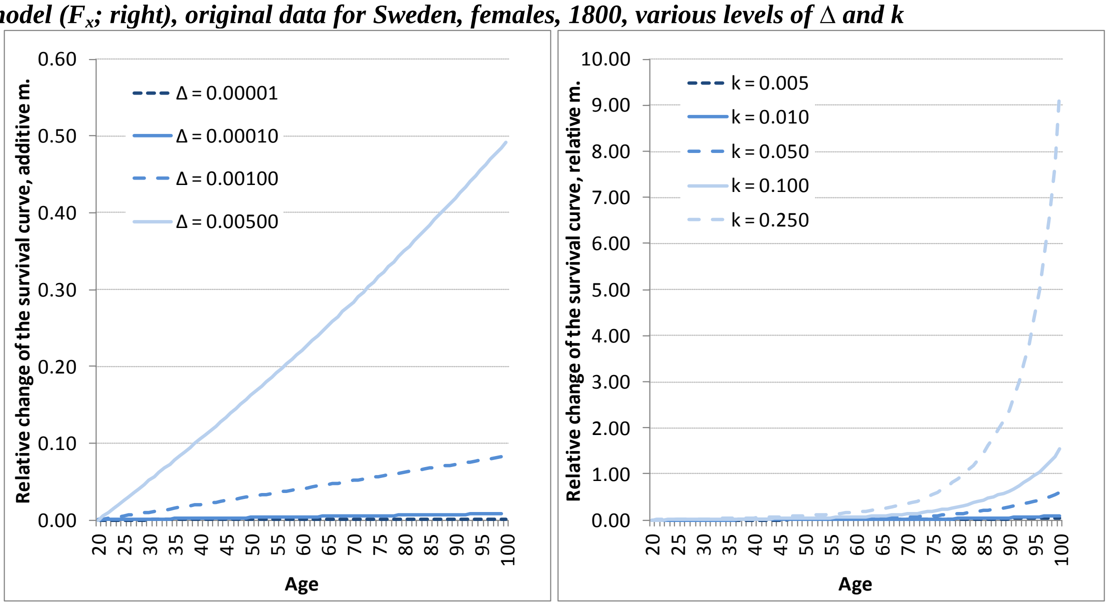
The results obtained for the model empirical data could be seen in the Figure 102.

Figure 102: Absolute and relative (R_x) change of the survival curve, relative model, k = 0.20, Sweden, females, 1800



Source of data: author's calculation based on Human Mortality Database (2010)

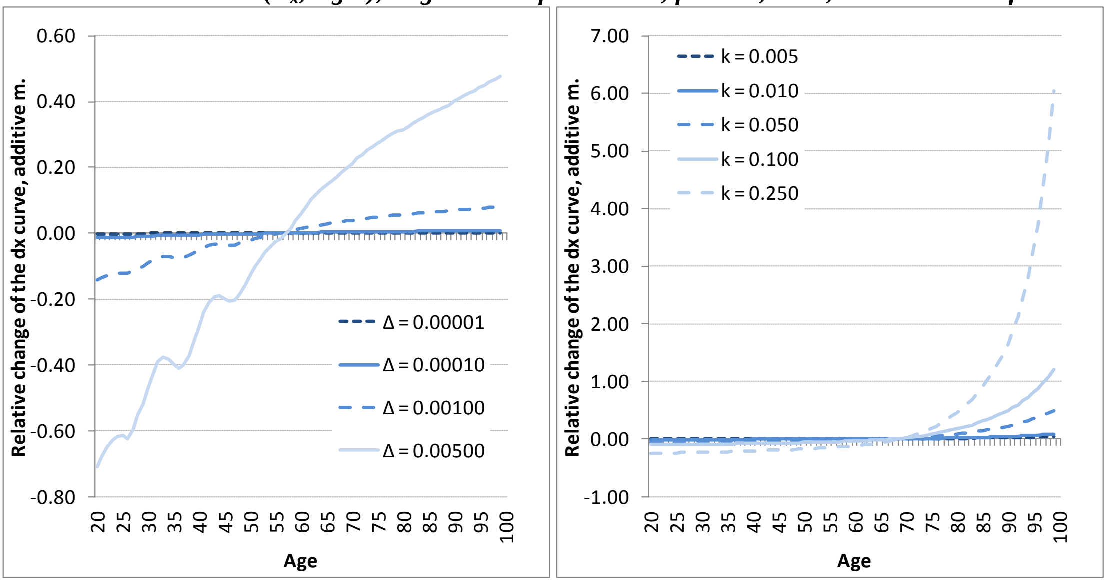
Figure 103: Relative change of the survival curve l_x in the additive model (φ_x; left) and in the relative model (F_x; right), original data for Sweden, females, 1800, various levels of Δ and k



Source of data: author's calculation based on Human Mortality Database (2010)

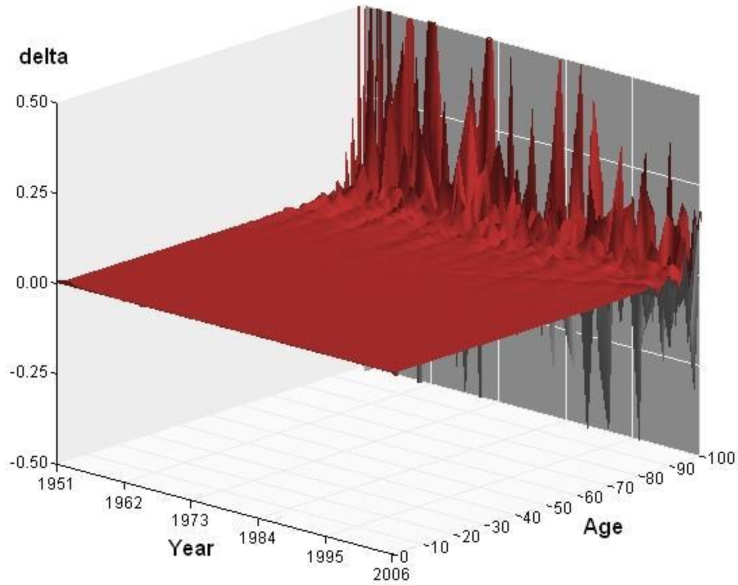
A simple simulation was done, in order to make it possible to see the differences of the relative changes of the table functions according to various values of the absolute change (Δ) or relative change (*k*) in mortality rates. At first, the data for Swedish females in the year 1800 were used together with the above derived formulas for the relative changes of the table functions in both the additive and relative model of mortality change. Then various values of the absolute change (Δ) or relative change (*k*) of mortality rates were imputed (Figures 103 and 104).

Figure 104: Relative change of the curve of the density of deaths, d_x, in the additive model (ρ_x; left) and in the relative model (R_x; right), original data for Sweden, females, 1800, various levels of Δ and k



Source of data: author's calculation based on Human Mortality Database (2010)

Figure 105: Absolute change of age-specific mortality rates (Δ_x), Czech Republic, males, 1950–2008

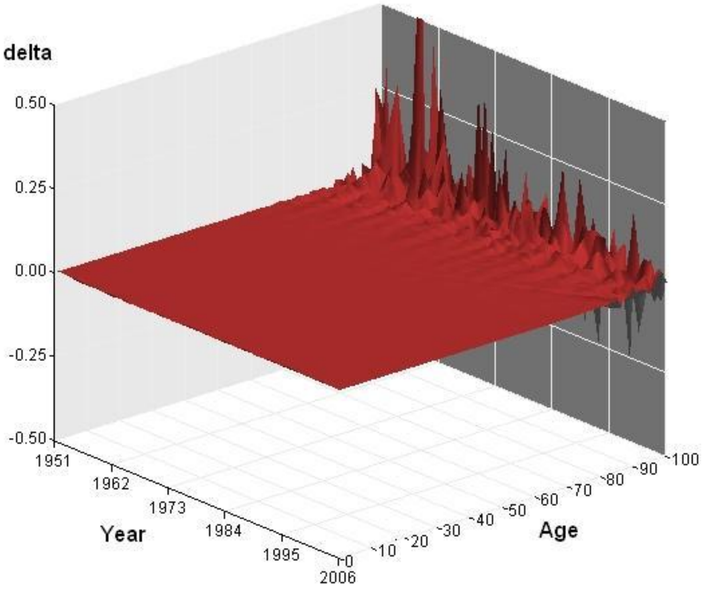


Note: Output from SAS 9.2 software

Source of data: author's calculation based on Human Mortality Database (2010)

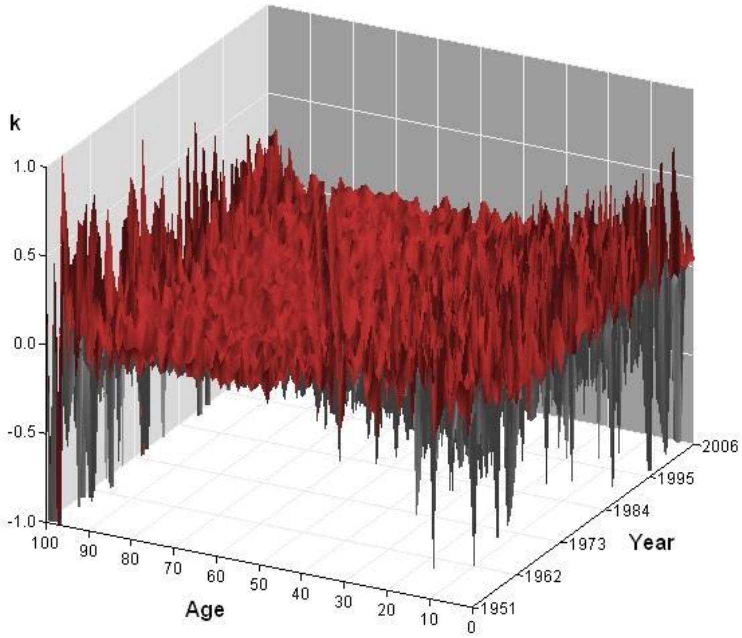
It could be seen that in general when there is a decrease of the age-specific mortality then as a result the survival curve increases above all at the higher ages. It is a consequence of higher proportion of survivors up to those high ages. The increase of the number of survivors at the highest ages is significant mainly in the situation of the relative decrease of the age-specific mortality rates, which means in a situation where the relative decrease is age-invariant.

Figure 106: Absolute change of age-specific mortality rates (Δ_x), Czech Republic, females, 1950–2008



Note: Output from SAS 9.2 software
Source of data: author's calculation based on Human Mortality Database (2010)

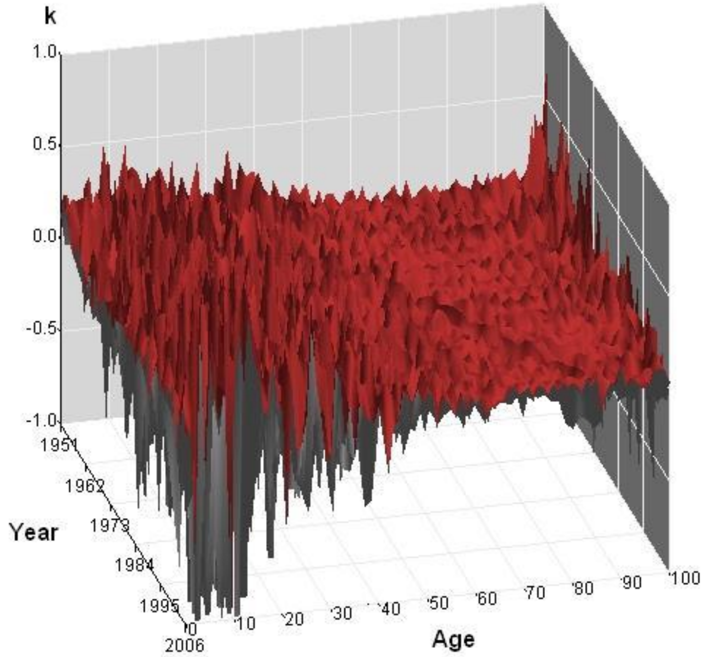
Figure 107: Relative change of age-specific mortality rates (k_x), Czech Republic, males, 1950–2008



Note: Output from SAS 9.2 software
Source of data: author's calculation based on Human Mortality Database (2010)

To find out which change of the age-specific mortality rates was more likely in the Czech Republic the absolute change and then also the relative change of mortality rates between each two following years from 1950 to 2008 was calculated for all ages. After a brief comparison between these absolute and relative changes it is clear that the relative change was not the same for all ages during the 2nd half of the 20th century. On the other hand the absolute change of the age-specific mortality is close to zero for all ages (except the highest ages) so the development of the year-to-year changes of age-specific mortality rates is closer to the nearly constant absolute change. That means (according to the Figures 105–108) that the increases of the number of survivors were relatively more uniformly distributed across ages in comparison to a hypothetical relative decrease constant across ages.

Figure 108: Relative change of age-specific mortality rates (k_x), Czech Republic, females, 1950–2008



Note: Output from SAS 9.2 software
Source of data: author's calculation based on Human Mortality Database (2010)

Based on the results above, in the following section, there the supposed relationships are developed also theoretically (without any empirical data) so as the formal relationships and assumptions of the mortality shifting could be seen. Again for simplicity the absolutely or relatively parallel change of the mortality rates was assumed.

8.6 Theoretical influence of the parallel decline/increase of the age specific mortality rates on the table functions – formal relationships

In the previous part of the text some basic equations illustrating the relationship between the parallel change (absolute or relative) of the age-specific mortality rates and its consequences for other life table functions were developed. In the figures, there the results were shown for a model data set corresponding with real Swedish data for females in 1800. Very important disadvantage of this data set was the fact, that the highest age where data were available was 100 years. There the table has to be ended (because no method of smoothing or extrapolation was used) no matter how big the shift was assumed. In the situation of significant decrease of the initial mortality rates the accumulation of deaths near the end of the life table occurred. As a consequence of that it was impossible to judge the situation fully and decide whether there appeared a shift which is measurable by the indicators of the mortality compression.

In this section other model data will be used – the hypothetical age-specific mortality rates corresponding to the logistic or Gompertz-Makeham functions. Again the age-specific mortality rates for ages under 20 were taken as being equal to zero. The highest attainable age was set as equal to 130, so that there are enough years where the curves could be shifted.

For the measuring of the rectangularization of the survival curve or compression of mortality some selected measures were chosen for the needs of this chapter. First of them is the modal age at death as a standard indicator used in mortality analysis. Then the interquartile range (IQR) is used as an indicator of the variability of ages at death. The third indicator used in the following analysis is the standard deviation of ages at death which occur above the modal age (SD+). The latest one describes the variability of ages at death at the highest ages – above the mode only. More about these indicators could be found in the Chapter 7.

When the studied population experiences mortality shifting (if it is understood as the horizontal movement of the survival curve or of the whole distribution of deaths to higher or lower ages without any change of its shape) we could expect the modal age to grow (or decline) while the standard deviation above the mode and the IQR should stay constant. Such a development of the selected measures would correspond to the movement of the life table functions above the age-axis. We will verify this assumption both for the additive and relative model (defined above) while the imputed initial mortality rates were generated from the logistic or the Gompertz-Makeham curve. For both of them several values of the absolute (Δ) and relative (*k*) decline in mortality rates will be applied. The aim of this part is then to find some basic general assumptions which have to be fulfilled so as the mortality shifting could be identified for a studied population.

At first we use the data generated by the logistic formula in the form (Thatcher, 1999; Thatcher *et al.*, 1998, with our symbolization):

$$\mu_{x,t} = \frac{A_t * \exp{(B_t * x)}}{1 + A_t * \exp{(B_t * x)}} + C_t \ .$$

Values of the parameters were chosen to be roughly corresponding to the average values of these parameters reached in the previous analysis in both groups of countries analyzed in the part devoted to the decomposition of the total mortality into the senescent and background components (Sub-chapter 8.3). In accordance to that the parameters in this part were considered as:

A = 0.00004
B = 0.12000
C = 0.00100.

Again we start with the application of the additive model to our data. The maximal value of the absolute decrease of initial mortality rates (Δ_{max}) was taken equal to the minimum of the mortality rates at ages 20 and more (for ages below 20 all the mortality rates were taken as equal to zero). In that case even the biggest possible parallel decline of mortality rates did not cause negative values of any of those rates. In our first model example Δ_{max} = *0.00144*. Several different values of Δ were then chosen, all of them were lower than Δ_{max}. The resulted changes of the survival curve could be seen in the Figure 109.

The absolute differences of the survival curve can be hardly seen in the picture. It is clear that the curve became more rectangularised because of its change at the young adult ages. In the picture below (Figure 110) it could be found that the relative change of the survival curve (φ_x) increased linearly with age but at higher ages the numbers of survivors are so low that even higher relative change of its value could not be significant in absolute terms.

Figure 109: Changes of the survival curve, additive model of mortality decreases, age-specific mortality rates generated by the logistic model, various selected Δ

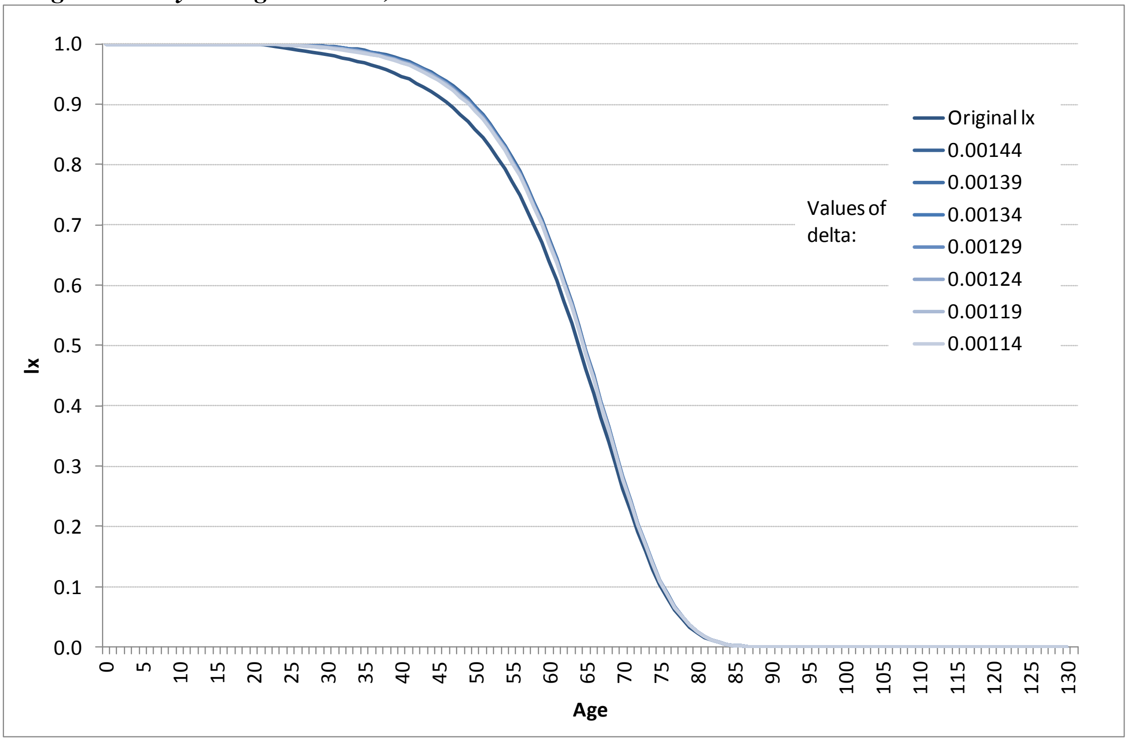


Table 14: Selected indicators of the mortality compression, additive model of mortality decreases, age-specific mortality rates generated by the logistic model, various selected Δ

delta (Δ)	IQR	mode	SD+
0.00144	13.27	66.78	17.28
0.00142	13.29	66.77	17.28
0.00139	13.31	66.77	17.28
0.00137	13.33	66.76	17.28
0.00134	13.35	66.76	17.28
0.00132	13.36	66.76	17.27
0.00129	13.38	66.75	17.27
0.00127	13.40	66.75	17.27
0.00124	13.42	66.75	17.27
0.00122	13.44	66.74	17.27
0.00119	13.45	66.74	17.27
0.00117	13.47	66.73	17.26
0.00114	13.49	66.73	17.26
0.00112	13.51	66.73	17.26
0 (original values)	14.37	66.56	17.18

Figure 110: Relative changes of the survival curve (φ_x), additive model of mortality decreases, age-specific mortality rates generated by the logistic model, various selected Δ

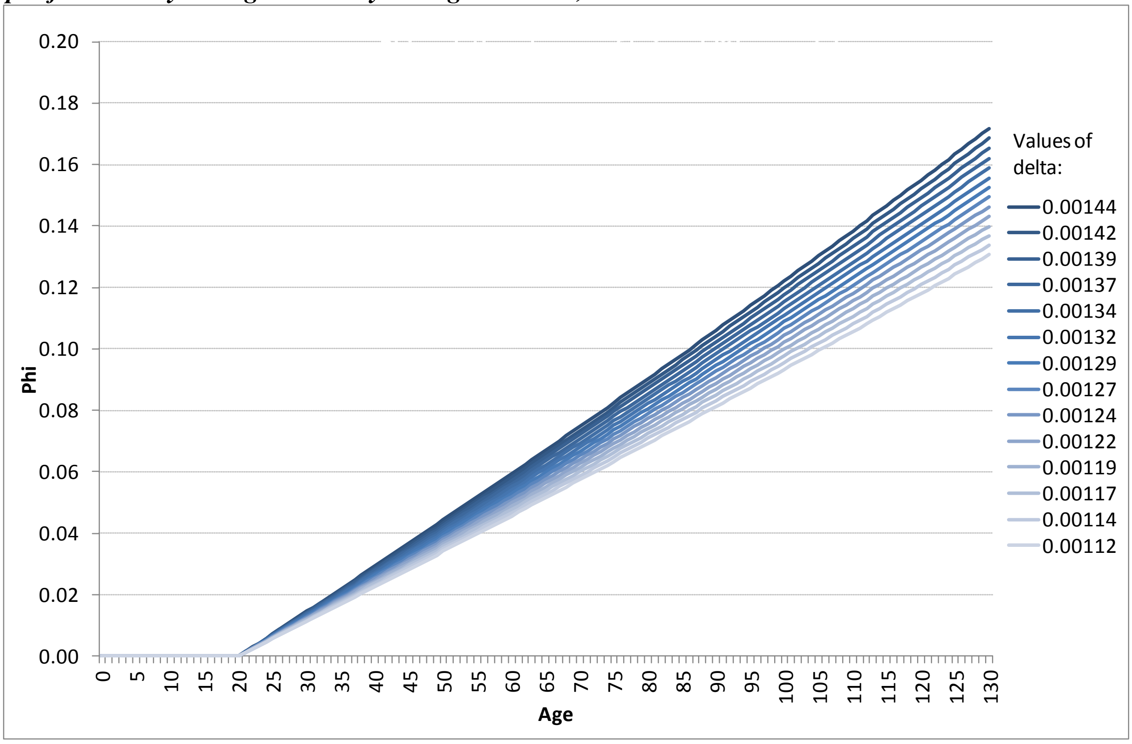
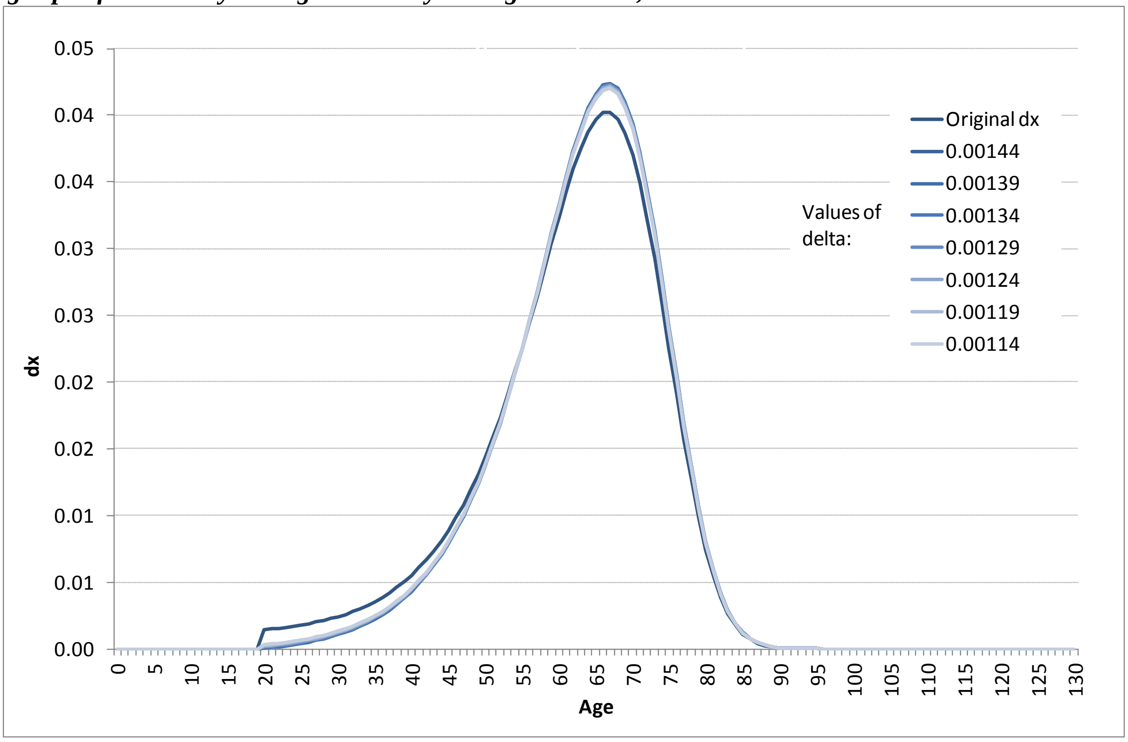
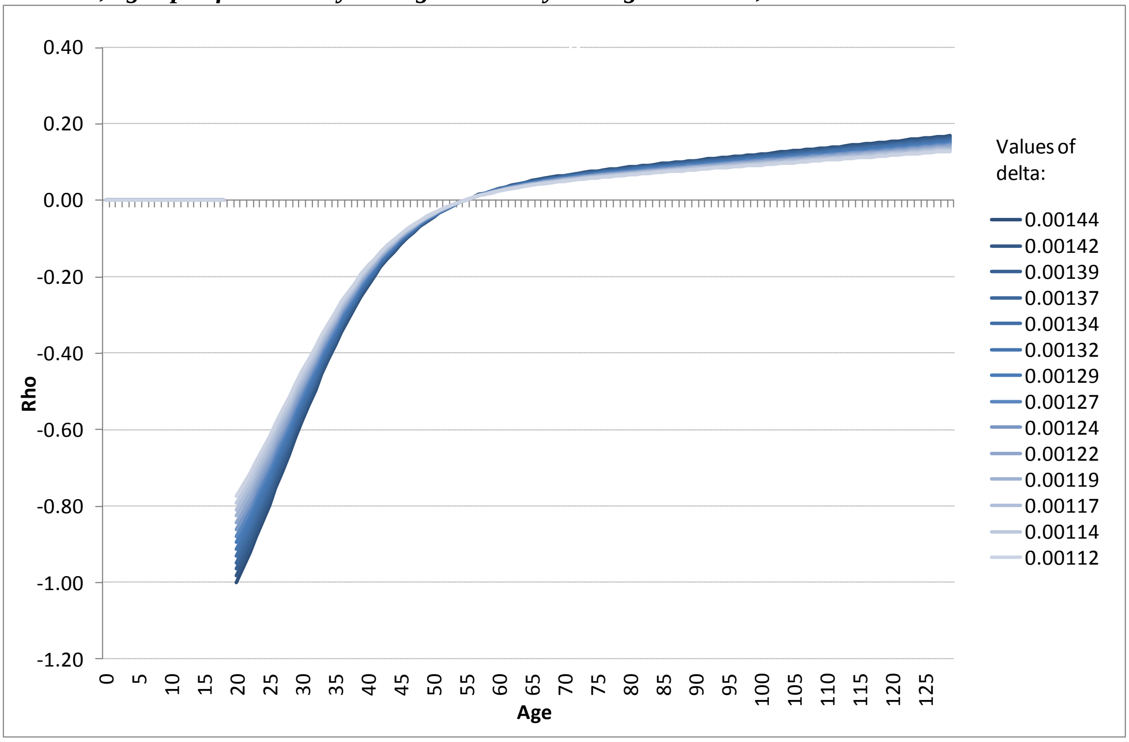


Figure 111: Changes of the curve of the distribution of deaths, additive model of mortality decreases, age-specific mortality rates generated by the logistic model, various selected Δ



From the selected indicators (Table 14), it could be seen that the mode was nearly stable during the time, it increased only slightly with the increase of Δ. Also the standard deviation of the ages at death above the mode (SD+) was almost stable. It also only slightly grew as a consequence of more survivors living at higher ages (above the mode) who also die at the ages above the mode. The only indicator which changed visibly was the IQR. So the variability of ages at death in the whole age interval decreased as a response to the supposed declines of mortality rates, but the variability above the mode (as was already stated) remained almost unchanged. That means that under the defined conditions the parallel absolute change of the mortality rates leads to the postponement of those deaths which otherwise would take part at lower ages. But if the variability of ages at death above the mode should stay almost unchanged, the number of deaths at the modal age has to increase. In the Figure 111 there it is proved that this happens in our model data.

Figure 112: Relative changes of the curve of the distribution of deaths (ρ_x), additive model of mortality decreases, age-specific mortality rates generated by the logistic model, various selected Δs



The above mentioned facts are supported by the Figure 112 where the relative change of the distribution of deaths under several defined conditions is depicted. The decreases at lower ages are relatively higher than the increases at higher ages. That signifies the higher concentration of deaths around the mode and the increase of the number of deaths which occur at the modal age.

The situation changes substantially when the relative model is used. Values of the relative decrease of age-specific mortality rates were selected to be minimally 5 % and maximally 99 % of the initial value. Curves for only selected values of *k* are depicted in the pictures.

What we can see in the Figure 113, it is closer to the standard notion of the shifts of table functions. The horizontal shifts of the survival curve seem to be almost parallel except for the last one (representing the highest decrease of mortality rates. So the higher the value of *k* (the relative decrease of the initial rates) the less parallel shift of the survival curve occurs. The reason for this is the fact that there the accumulation of the survivors to the highest ages is present. That is also proved by the values of the relative change of the survival curve (F_x) in the Figure 114. The bigger the decline of the initial rates, the more rapidly the relative change of the survival curve rises with age.

Figure 113: Changes of the survival curve, relative model of mortality decreases, age-specific mortality rates generated by the logistic model, various selected k

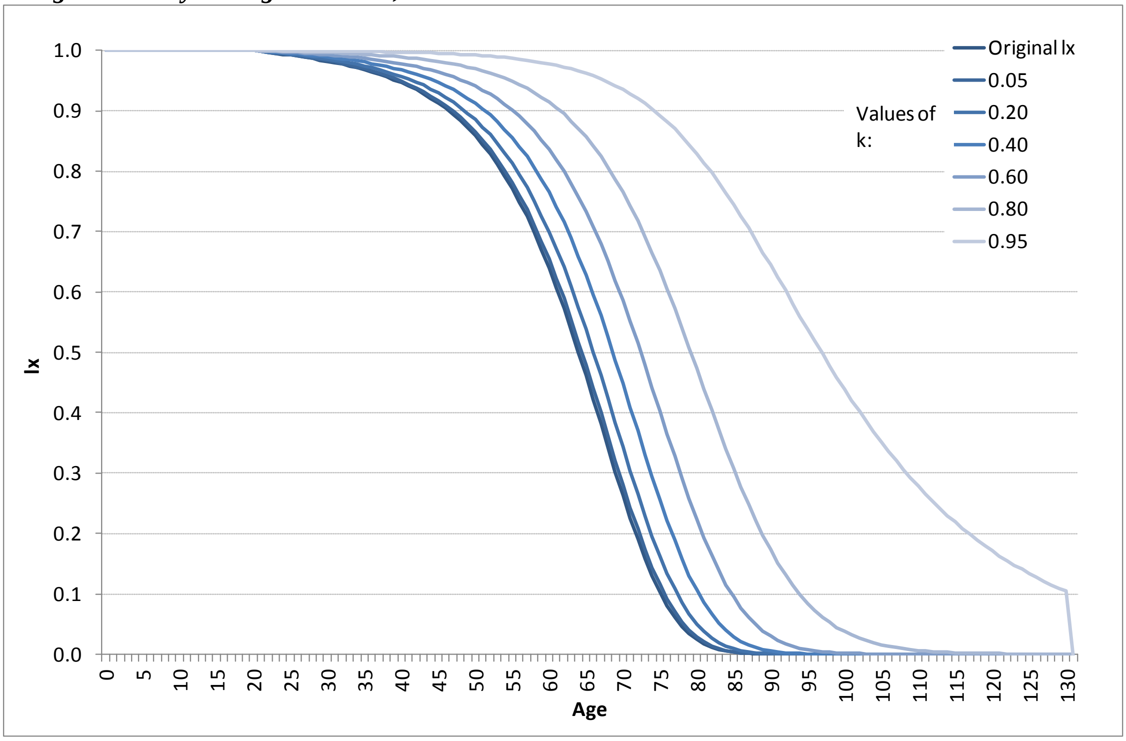
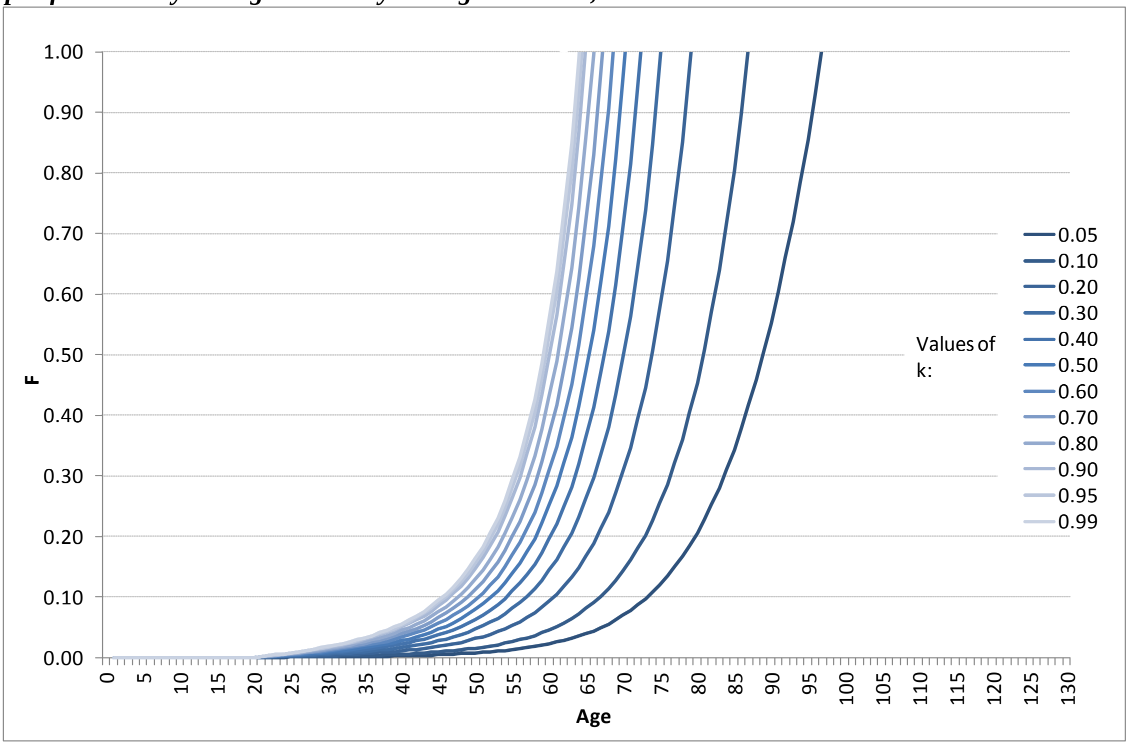


Figure 114: Relative changes of the survival curve (F_x), relative model of mortality decreases, age-specific mortality rates generated by the logistic model, various selected k



In the case of the relative model the mode is increasing with the values of *k*. The tempo of the increase seems to be higher for higher values of *k*. The rising values of the mode could be the proof of the shifting in data. But it is not. The reason is clear; the standard deviation of ages at death above the mode is growing also. And also the IQR, as a representative of the total variability of ages at death, is increasing with the values of *k*. But it could be seen in the

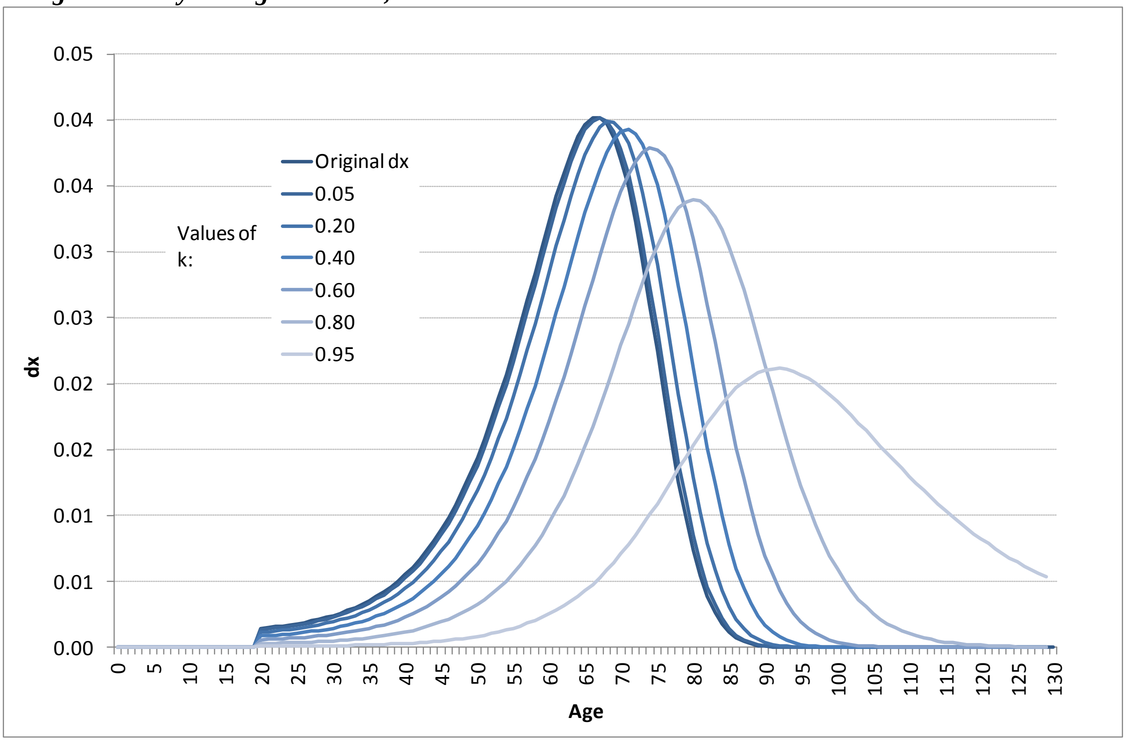
Table 15 that for *k* <*0.6* the changes of the IQR are not significant and values of the standard deviation above the mode are almost constant for *k* <*0.3* while the mode is increasing for all *k*.

Table 15: Selected indicators of the mortality compression, relative model of mortality decreases, age-specific mortality rates generated by the logistic model, various selected k

k	IQR	mode	SD+
0 (original values)	14.37	66.56	17.18
0.05	14.38	66.98	18.18
0.10	14.38	67.45	17.56
0.20	14.41	68.44	18.29
0.30	14.46	69.57	19.49
0.40	14.53	70.86	21.41
0.50	14.66	72.40	22.64
0.60	14.89	74.28	25.78
0.70	15.31	76.69	32.98
0.80	16.28	80.08	47.26
0.90	19.60	85.88	91.15
0.95	27.50	91.67	102.96
0.99	77.32	105.10	–

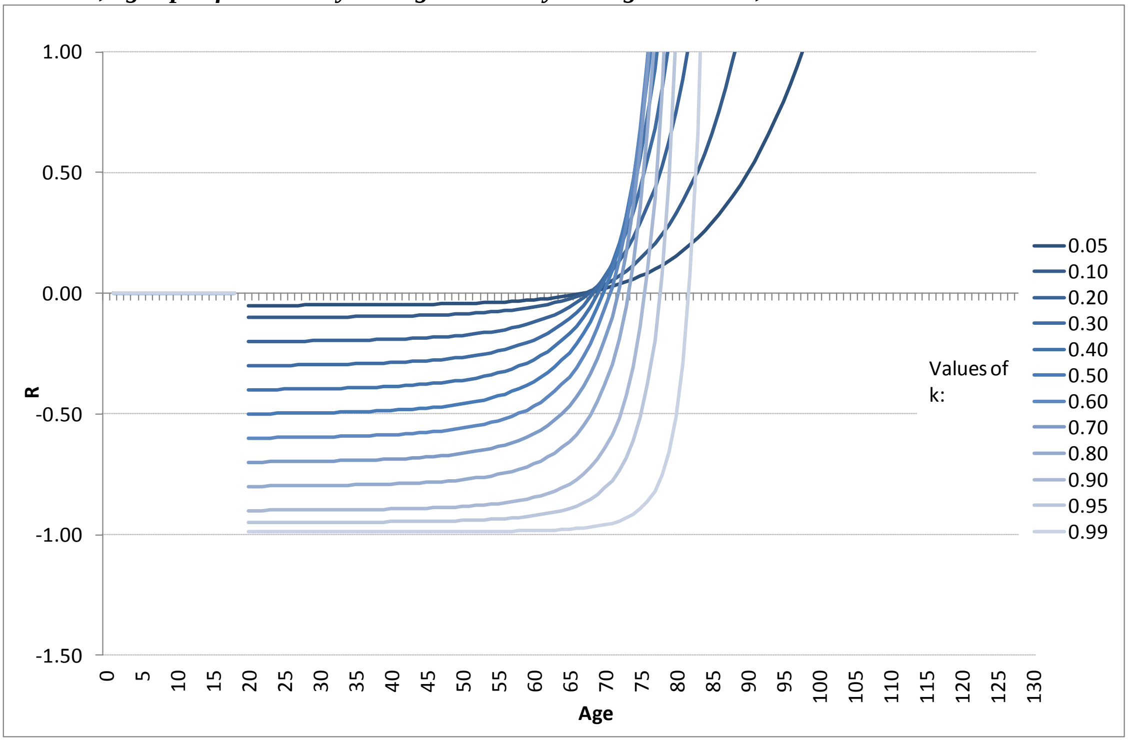
Therefore, with some extend of simplification, it could be concluded that for logistically increasing mortality rates where the relative parallel decline occurs the result is close to the horizontal shifting of the survival curve and of the distribution of deaths. The shifts of the deaths curve could be seen in the Figure 115. It is confirmed that for *k* higher than ca *0.3* the process could not be called "shifting" because the rising variability of the ages at death becomes more important with rising *k* and as was stated above, the stability of the variability of ages at death was taken as a condition for the shifting of the table functions.

Figure 115: Changes of the curve of deaths, relative model of mortality decreases, age-specific mortality rates generated by the logistic model, various selected k



All the said is sketched in the Figure 116 where the relative change of the curve of deaths is presented for several values of *k*.

Figure 116: Relative changes of the curve of distribution of deaths (R_x), relative model of mortality decreases, age-specific mortality rates generated by the logistic model, various selected k



Now the previous analysis will be repeated but for different initial conditions. For this second example, we suppose the age-specific mortality rates to follow the Gompertz-Makeham function defined as

$$\mu_{x,t} = A_t + B_t * \exp(C_t * x),$$

where the parameter values were taken in accordance to the actual situation of females in the Czech Republic, because the Czech Republic could be taken as a representative of almost an average country (according to the level of mortality) in Europe – it is one of the best ones in the post-communist block and one of the slightly worse ones in comparison to the non-post-communist countries in Europe. For our model the parameters were taken as:

A = *0.000286000*
B = *0.000005507*
C = *0.116365000*.

For completeness it should be repeated that again all the mortality rates under age 20 were taken as being equal to zero and no further decrease was applied to them.

In the additive model the value of the maximal absolute decrease of the rates (Δ_{max}) was calculated in the same manner as in the previous case – it was taken as to be equal to the minimal mortality rate for the ages 20 and more. As a result the Δ_{max} was taken as *0.00034* and the lowest Δ incorporated in the simulation was equal to *0.00002*. All the three measures (mode, IQR, SD+) were used also in this example.

Figure 117: Changes of the survival curve, additive model of mortality decreases, age-specific mortality rates generated by the exponential (Gompertz-Makeham) model, various selected Δ

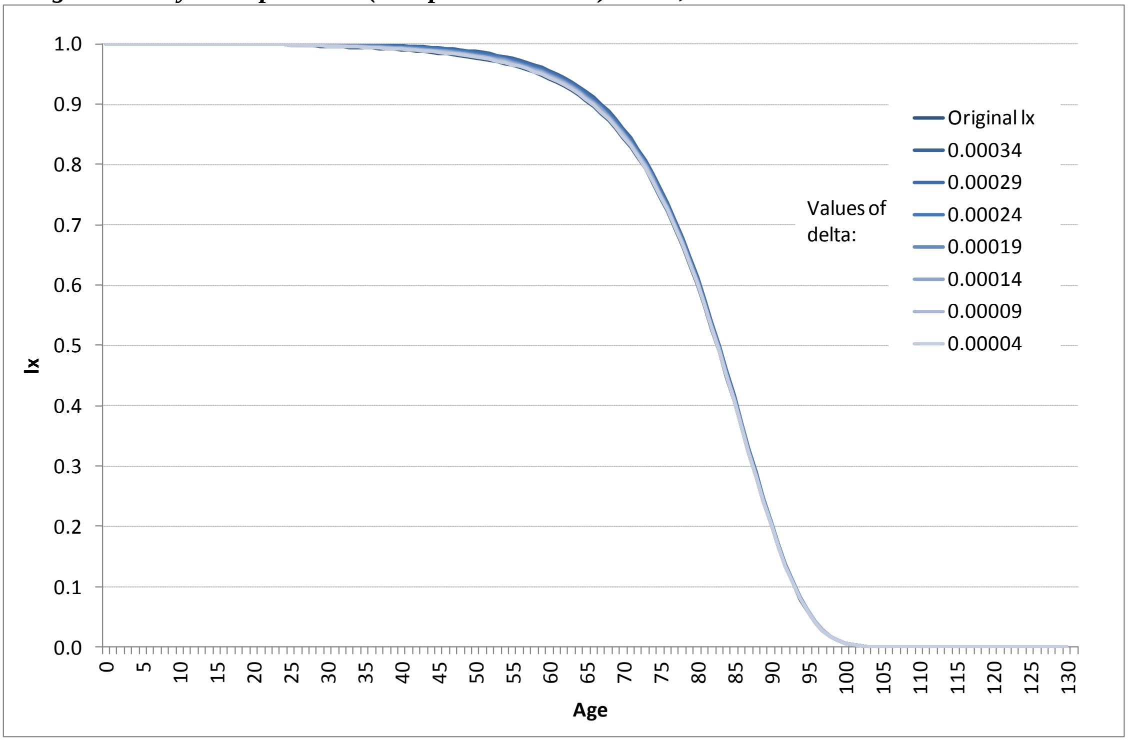
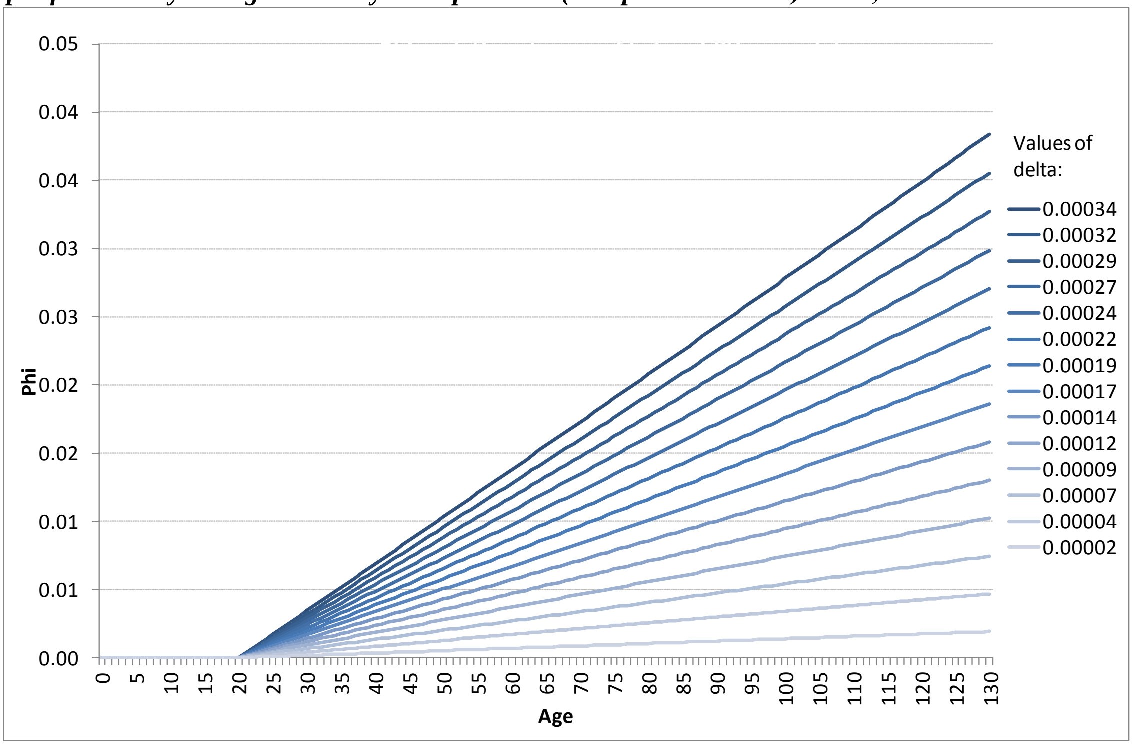


Figure 118: Relative changes of the survival curve (φ_x), additive model of mortality decreases, age-specific mortality rates generated by the exponential (Gompertz-Makeham) model, various selected Δ



First of all the additive model was applied to the data. The look of the changes of the survival curve (Figure 117) seems to be similar to the case where the logistic model for mortality rates was considered. This fact is supported also by the values of the relative change of the survival curve in the Figure 118. The pattern is the same but the values differ. The logical conclusion could be that in the additive model of the parallel mortality decline, there the shape of the initial mortality curve (the distribution of age-specific mortality rates with age) is not important. There the changes resulting from that absolute parallel decline/increase of mortality rates lead always to the same pattern of the change of the table functions (l_x, d_x). Again this is confirmed by the figures where changes of the curve of distribution of deaths could be seen (Figures 119, 120).

Moreover it is supported also by the values of the indicators used here (Table 16). The mode and the IQR are almost constant during the decline, the only indicator which shows some change is the SD+. The standard deviation of the ages at death above the mode is increasing with Δ as a result of the higher accumulation of survivors (and deaths) above the modal age at death.

Figure 119: Changes of the curve of the distribution of deaths, additive model of mortality decreases, age-specific mortality rates generated by the exponential (Gompertz-Makeham) model, various selected Δ

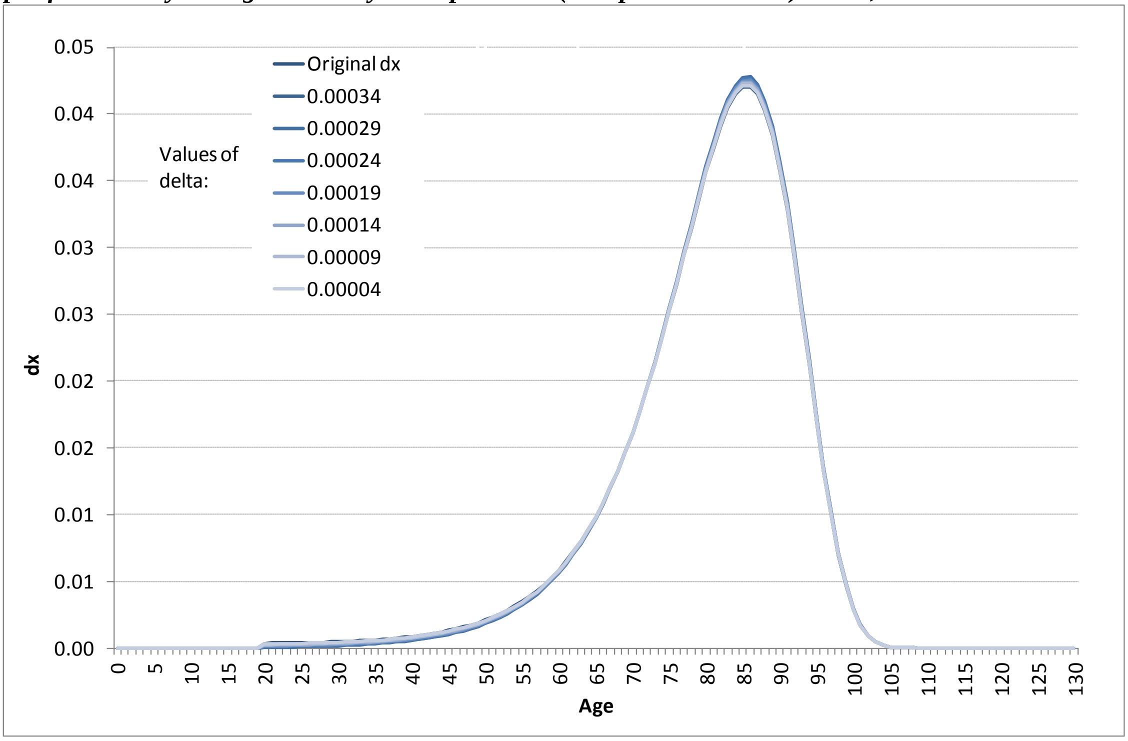


Figure 120: Relative changes of the deaths curve (ρ_x), additive model of mortality decreases, age-specific mortality rates generated by the exponential (Gompertz-Makeham) model, various selected Δ

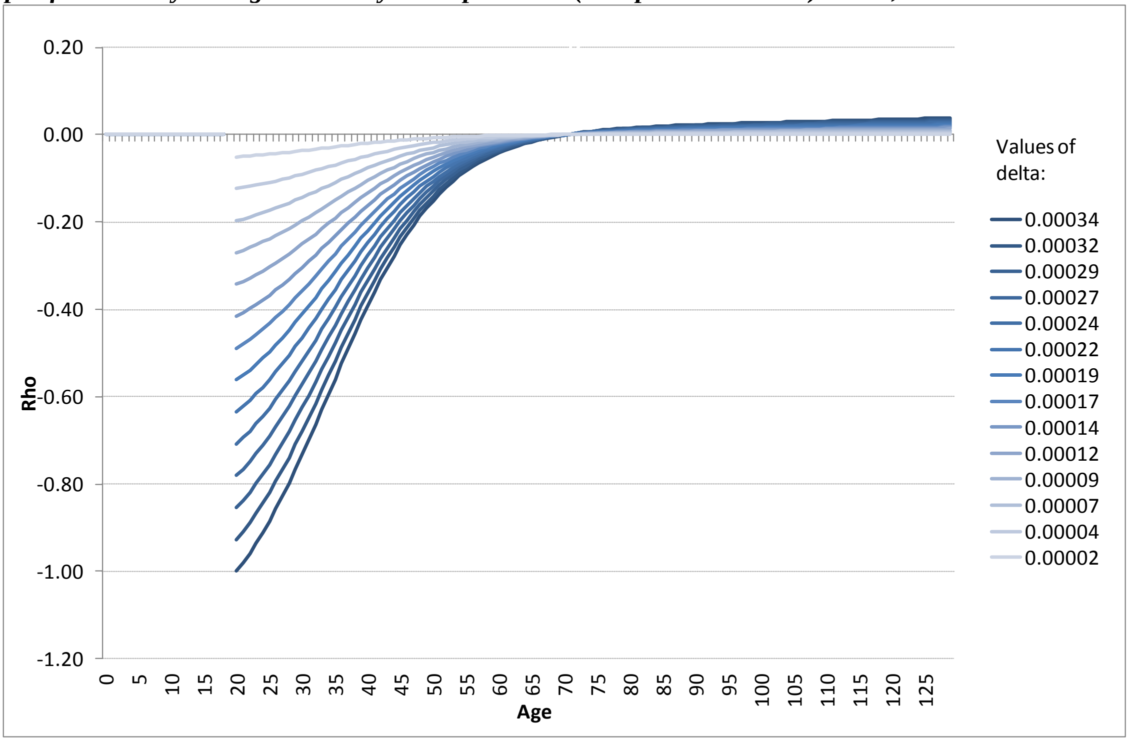


Table 16: Selected indicators of the mortality compression, additive model of mortality decreases, age-specific mortality rates generated by the exponential (Gompertz-Makeham) model, various selected Δ

delta	IQR	mode	SD+
0.00034	13.45	85.58	13.20
0.00032	13.48	85.58	11.44
0.00029	13.51	85.57	11.44
0.00027	13.54	85.57	11.43
0.00024	13.57	85.57	11.43
0.00022	13.60	85.56	11.43
0.00019	13.63	85.56	11.43
0.00017	13.66	85.55	11.43
0.00014	13.69	85.55	11.43
0.00012	13.72	85.55	11.43
0.00009	13.75	85.54	11.43
0.00007	13.78	85.54	11.43
0.00004	13.81	85.54	11.43
0.00002	13.85	85.53	11.43
0 (original values)	13.87	85.53	11.43

Finally we will consider the exponentially (following the Gompertz-Makeham function) increasing mortality rates in the relative model. Values of the relative decline were taken the same as in the previous example where the logistically increasing mortality rates were imputed into the relative model.

Figure 121: Changes of the survival curve, relative model of mortality decreases, age-specific mortality rates generated by the exponential (Gompertz-Makeham) model, various selected k

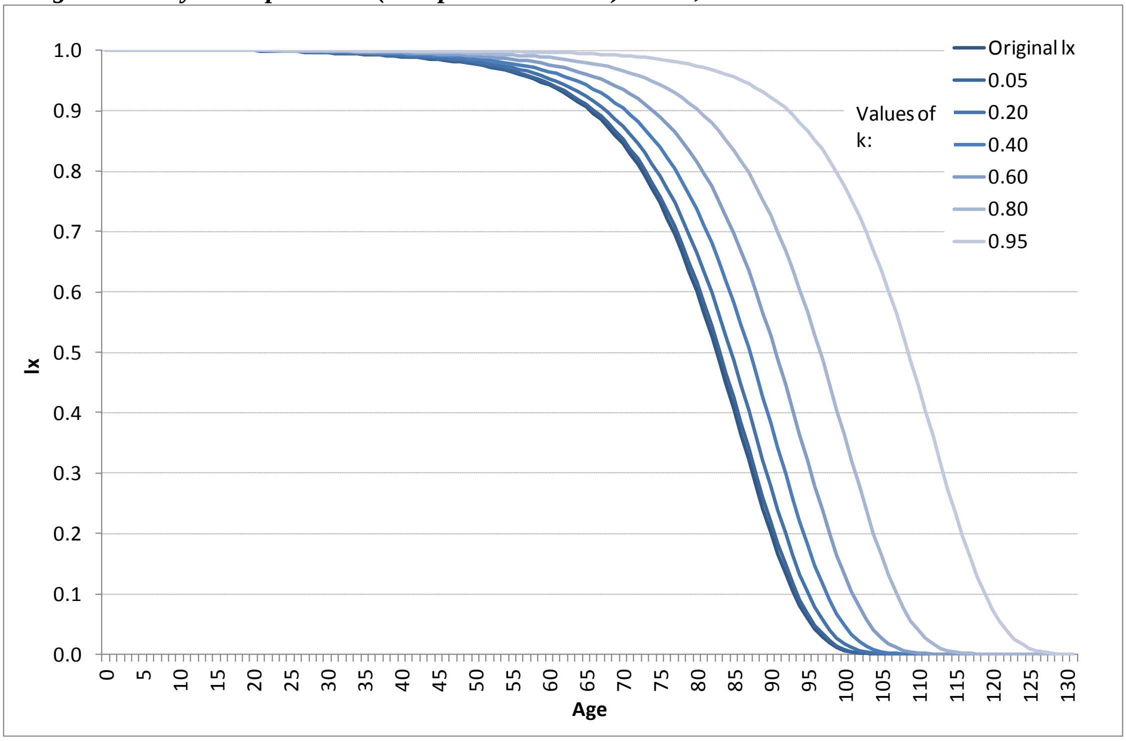
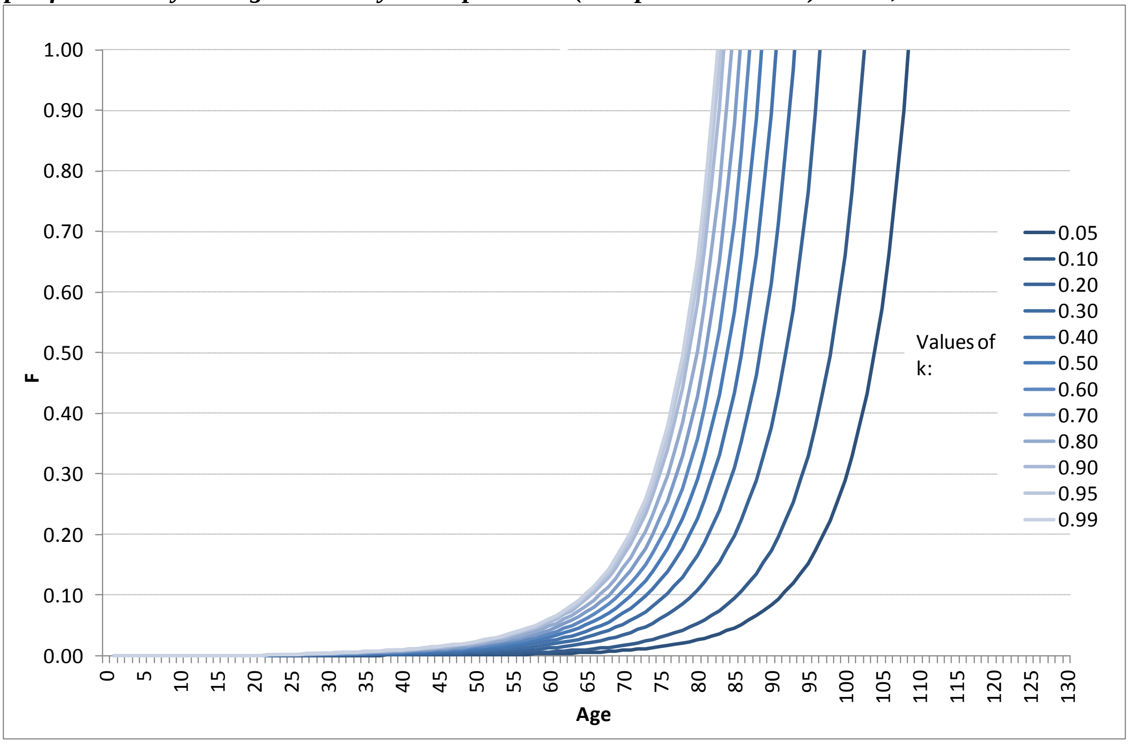


Figure 122: Relative changes of the survival curve (F_x), relative model of mortality decreases, age-specific mortality rates generated by the exponential (Gompertz-Makeham) model, various selected k



What could be seen in the Figure 121 (and is confirmed by the Figure 122) it is the real shifting of the survival curve. The horizontal differences between each two curves are identical for all ages and this holds for all possible values of the relative decline of the initial mortality rates, that means for all *k*.

Table 17: Selected indicators of the mortality compression, relative model of mortality decreases, age-specific mortality rates generated by the exponential (Gompertz-Makeham) model, various selected k

k	IQR	mode	SD+
0 (original values)	13.87	85.53	13.19
0.05	13.86	85.96	13.82
0.10	13.84	86.44	12.98
0.20	13.81	87.46	13.00
0.30	13.78	88.61	13.23
0.40	13.75	89.93	13.75
0.50	13.71	91.51	13.07
0.60	13.68	93.43	12.94
0.70	13.65	95.89	13.68
0.80	13.61	99.39	12.88
0.90	13.57	105.35	12.81
0.95	13.55	111.31	12.62
0.99	13.53	125.14	1.24

The fact that for the exponentially increasing intensity of mortality, where we suppose the relative parallel shift of its values, we can get the "real" shifting of the table functions is confirmed also by the values of our selected indicators (Table 17). The modal age at death is increasing with the values of *k* while the other two indicators remain almost constant (except for the extreme values of *k* where the reason is the ending of the life table where survivors and deaths are more cumulated near the maximal age of the table (age 130 in our case).

The shifting of the table functions is shown also in the Figures 123 and 124 where the changes of the curve of distribution of deaths could be studied in more detail.

Figure 123: Changes of the curve of the distribution of deaths, relative model of mortality decreases, age-specific mortality rates generated by the exponential (Gompertz-Makeham) model, various selected k

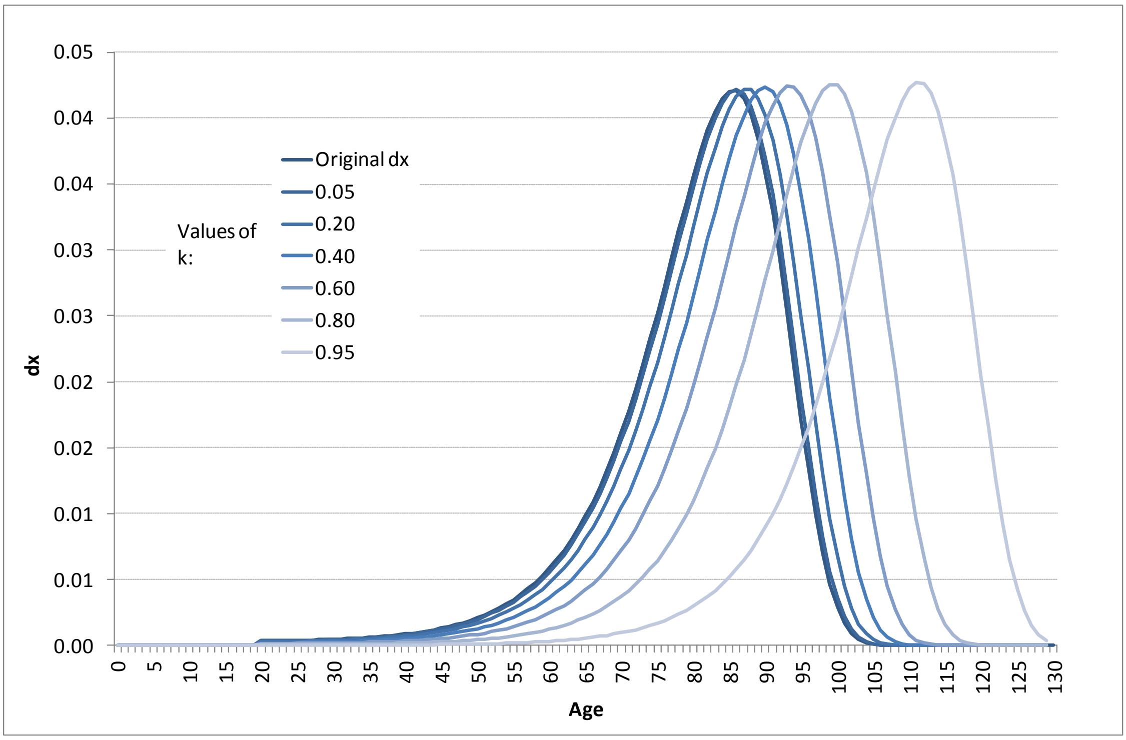
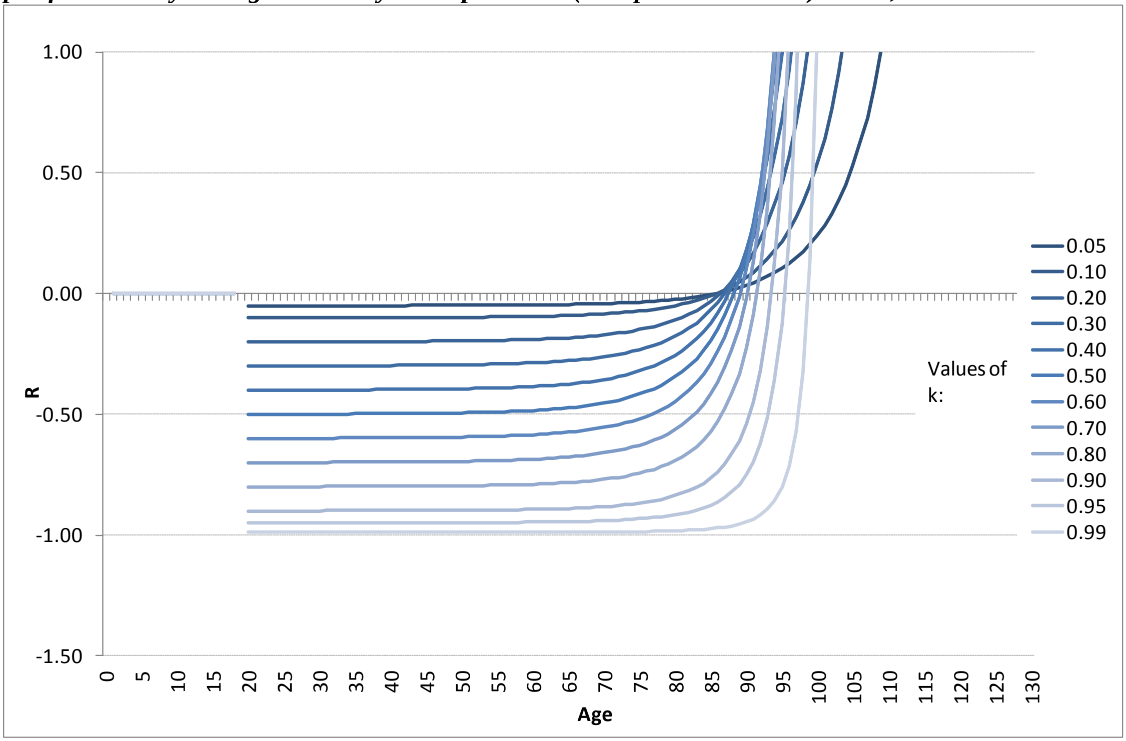


Figure 124: Relative changes of the death curve (R_x), relative model of mortality decreases, age-specific mortality rates generated by the exponential (Gompertz-Makeham) model, various selected k



It could be concluded that there were two clear conditions for connection of the parallel change of the mortality rates and parallel horizontal shift of the life table functions (in other words for the appearance of the horizontal parallel shift of the table functions) described in the text above. First, the "real" shifting of the table functions could be seen only in that case when the relative parallel decrease of the initial mortality rates occurs, that means that the relative decrease of the initial mortality rates does not change with age. The percentage of such a decrease was marked by *k*. And second, the initial mortality rates have to be exponentially increasing with age. In our example the Gompertz-Makeham function was used.

8.7 Conclusions

The theme of this chapter may seem to be divided into several almost independent issues (the decomposition of mortality into its two components, study of the senescent mortality, morality shifting and its assumptions), but all those issues are so closely related that it was the reason for putting them all into one part of the Thesis. At the very beginning of this chapter the main idea and concept of the distinguishing between background and senescent mortality was introduced. This separation of both the components enables the study of the mortality development in more detail. It was concluded that the stagnation of mortality decline at the beginning of the second half of the 20th century in many developed countries was caused by the stabilization of the background mortality (the component invariant with age) on already very low values. The following mortality decline, seen above all at the end of the 20th century and at the beginning of the 21st century, is the consequence of decreasing values of the senescent mortality (the component dependent on age). Both these mortality components could be separated by many methods, however, within this chapter two mortality functions (Gompertz-Makeham and Thatcher) were used. After the estimation of their parameters (by the method of the least weighted non-linear squares) both the mortality components could be easily calculated. Then the main differences between males and females or between two different groups of countries were studied. For this purpose all the countries, where data were available, were divided into post-communist and other countries. It showed out that the development in the post-communist countries is in many aspects similar to the development of the other, mostly Western or Northern European, countries. There is only some time lag existing which could be seen for example in the figures above. The exception of this pattern were some mostly post-soviet countries where the mortality development at the end of the 20th century and during the last decades was so unfavorable that almost no clear developmental trends similar to the non-post-communist countries could be traced. This mortality development was the consequence of many socio-economical problems arising from the change of the political regime.

The senescent component could be separately studied in deeper analysis. Bongaarts (2005) when dealing with this issue proposed to take one of the parameters of the logistic curve as a constant and he derived the formal relations connected with the process of mortality shifting. However, this assumption was evaluated before it was applied also to the data used in the analysis within this Thesis. It showed that the mentioned condition could be taken as verified only for some contemporary or low-mortality countries. However, within this Thesis more countries with various mortality levels were analyzed and where it was possible also the historical data were used. In that case it did not show to be appropriate to keep the Bongaarts' assumption. The instability of the parameter (taken by Bongaarts as a constant) was proved also by the development of "age-specific shifts" as it was called in the text. Through this tool it could be concluded that in the studied population the pure horizontal mortality shifting occurs only in such a situation when the age-specific shifts are constant with age.

Finally, the shifting process was studied also in a more theoretical, or formal, way. For that purpose several formal relations were derived which describe the behavior of two selected table functions (the survival function and density function) in a response to a change in the age-specific mortality rates. This change in mortality rates was for simplicity taken as a constant absolute decrease/increase or constant relative decrease/increase. First of all the formal relations were illustrated on real data (historical data for Sweden) and then the same relations were applied to model data generated according the exponential Gompertz-Makeham and logistic Thatcher functions. It was shown that the pure mortality shifting (understood as a horizontal shift of the table function without any change of the shape of their curves) could occur for exponentially increasing force of mortality where the relatively constant decline of mortality (the same percentage decline for all the ages) appears. For lower values of the percentage decline of mortality rates it was concluded that the shape of the initial intensity of morality (exponential or logistic) does not play so significant role in comparison to the difference between the absolute or relative decline. For the Czech data it was shown that in the second half of the 20th century and at the beginning of the 21st century the declines of the age-specific mortality rates are closer to the absolute decline (similar value of the decline for almost all the considered ages, this change is close to zero between two subsequent years) then to the relatively constant decline (because the relative declines in the Czech population between each two subsequent years was highly dependent on age).

All the analyses confirmed mainly the similar developmental trends in post-communist countries (except the special case of the post-soviet countries) as could be seen in the Western or Northern Europe. The information about the time lag of these trends could be used not only for the interpretation of similar analyses but also for forecasts of the future development. The same could be said about the results of the decomposition of the total mortality to its background and senescent component because it is possible to study both those components separately. The mortality shifting, a special form of the mortality development, could be seen only in some generally low-mortality countries during the last years. It was shown, however, how this process could be identified in the data set. When a pure mortality shifting occurs in the data and when it is identified, it could also simplify the process of mortality forecasting because in that situation some relatively stable formal relations could be used.

Wrinkles should merely indicate
where smiles have been

Mark Twain, Following the Equator

Chapter 9

The role of tempo effect in mortality analysis

Demographers are since the time of Graunt, Halley and others used to use the period mortality rates for construction of the life tables which, as was said already, are one of the most important and most universally used tools in demography. Usually there was no reason to think about those methods in a more critical manner. The so-called tempo effect and its consequences for traditionally used measures is still a bit controversial theme in demography (Vaupel, 2009; Barbi *et al.*, 2008; Luy, 2010a; Luy, 2010b). This theme was included to this Thesis for several reasons. First of all the analysis of tempo effects is related to the issue of life tables and period indicators in general. Those topics are the base of the whole Thesis and the first part of the text was almost entirely devoted to the life tables. The second reason is the fact that the tempo effects and their analysis are also partially connected with the transition of the mortality analysis to the usage of cohort measures. For the same reason also the 10th chapter devoted to frailty models is incorporated into the Thesis. It has to be mentioned also that the issue of tempo effect is so interesting and discussed that there is worth dealing with that. Moreover the application of some tempo-adjusting methods introduced later can significantly change the previous results.

Within this chapter the role of tempo effects will be introduced briefly together with some theoretical illustrations. Then the most important methods of tempo-adjustment will be shown and their main assumptions will be verified. Finally results obtained by different methods will be compared. It will be proven that the results could differ significantly in some cases.

9.1 Introduction and theoretical background

The first interests in some distortion of traditional measures were inspired by the effort to evaluate the relationship between the numbers of births and the changes in timing of fertility represented usually by the mean age at birth (Ryder, 1956; Bongaarts, Feeney, 2003). Ryder (1956) introduced also the term “timing distortion”. The first process where the distortion was analyzed was fertility (also by Ryder, 1956) but particular methods for tempo adjusting were proposed many years later (in 1998 by Bongaarts and Feeney). In 2002 John Bongaarts and Griffith Feeney introduced the idea that the same type of distortions could be observable also for mortality, when the rates are constructed in the same manner as they are in fertility analysis

(Bongaarts, Feeney, 2002). They noted that the distortion affects the numerators of all period event rates (Barbi *et al.*, 2008). On the basis of the previous research, the authors assumed that the growth of the mean age at death could lead to the fact that the period life expectancy is overestimated and that it is underestimated when the mean age at death is decreasing. They argue that this would also distort the trends of life expectancy which are used in mortality analysis and in population forecasts (Bongaarts, Feeney, 2002). The distortion is a consequence of the fact that the period life expectancy in time t depends solely on the force of mortality function for time t. The adjusted measure, proposed by Bongaarts and Feeney and described below depends not only on the force of mortality in the studied year, but also on the rate of change of the mean age at death (Bongaarts, Feeney, 2008). Some methods of tempo adjustment are already practiced within fertility analysis but the mortality tempo adjustment is less accepted and used in the practice of mortality analysis so far (Luy, 2010b).

One could argue that we can avoid these possible distortions by using the cohort measure. Of course, the cohort life expectancy could not be distorted at all if we exclude the possibility of distorted input data. But the important disadvantages of such measures are clear. First of all long time series of historical data are needed for cohort indicators. Only for a limited number of countries such data are available (e.g. in the Human Mortality Database). In some countries, what is also the case of the Czech Republic, the time series of relevant data is interrupted by the periods of World Wars for which the data are not existing or not enough reliable. Except for data, there are other disadvantages of measures based on cohort life tables: the cohort life expectancy represents in fact a summary of mortality of more decades, usually a century or even more. During that time the conditions influencing mortality are likely to have changed significantly (Bongaarts, Feeney, 2002). In that case even the interpretation of such a measure is already quite problematic. In addition to that, the cohort life expectancy does not allow the study of some temporary changes in more detail.

The disadvantages mentioned above and the lack of cohort data were the main reasons why for the study of mortality trends the period life expectancy has become one of the most frequently used tool. The period life expectancy could be (and often it is) interpreted as “the average age at death that would be observed for a group of persons who experience, over the course of their lives, the age-specific death rates observed during the time period” (Bongaarts, Feeney, 2002, p. 14). This might be taken as acceptable way of interpretation. But what happens if we incorporate also the fact of changing timing of events, which means the timing of deaths in the case of mortality? What does it mean for the analysis if we suppose, for example, the continuing process of mortality shifting (introduced in previous parts of the Thesis) assuming that we agree upon a definition and presence of it?

Luy defined four typical features of conventional period indicators; those are (Luy, 2010b, p. 421):

1) period indicators represent the demographic conditions of the current studied year (or period of years);
2) period indicators have usually clear, interpretable and easily understandable meaning;

3) period indicators refer to hypothetical populations;
4) period indicators are standardized for age.

Luy (2010b) than supposes that the idea of Bongaarts and Feeney is simply the extension of the 4th feature mentioned above. The other three features hold also for the tempo-adjusted indicators. Then the "tempo effect" could be defined as "a change of period rates for demographic events […] that solely results from a change of the average age at which the event occurs during the observation period" (Luy, 2010b, p. 421; in other words also Yi, Land, 2002).

Bongaarts and Feeney (2002) present an illustrative example of the tempo effect and its consequences. They supposed a stationary population where during one period (one year) the mean age at death linearly increased. Then they showed that the life expectancy calculated using the standard life table methods will increase sharply as a result of the tempo effect leading to temporary decline of numbers of deaths.

The example shown in the articles of Bongaarts and Feeney from 2003 and 2006 (Bongaarts, Feeney, 2003; 2006) is probably more illustrative and clear. Again they suppose a stationary population with the life expectancy equal to 70 years where a sudden change happens in a year *T*, when all the deaths are postponed for some time. We can imagine that development, in accordance to Bongaarts and Feeney, as a consequence of a production of something like a "life extension pill" which each person in the studied population takes at the beginning of the year *T* (*ibid*.). This pill leads to a life extension (postponement of death) for, say, 3 months. From that time of the pill's application nobody would die during 3 months. That means that the number of deaths in the year *T* will decrease by 25 % (*ibid*.). The authors did not emphasize one very important assumption for the rest of this example – that the application of the "life extension pill" will only shift the mortality curve to higher age. Therefore the survivors after three months from the application of the pill will die according to exactly the same mortality curve which is now only shifted to the age higher for three months. In other words it means that these people whose life was extended or even saved by the pill would not have completely different level of mortality. This assumption is mentioned by Bongaarts and Feeney in the general description of the model, it comes more clear, however, from the articles of Vaupel (2002; 2009) and Vaupel and Yashin (1986 and 1987), who introduced the idea of different mortality regime of the "resuscitated" (Vaupel, Yashin, 1986; 1987). But, in this moment, we do not need these models connected also with the issue of heterogeneity of population (the population heterogeneity is the theme of the next chapter). Let's suppose the pill which does not change the mortality curve but leads only to its shift to higher ages.

Bongaarts and Feeney (2003; 2006) showed that when we suppose the 100 % effectiveness of the pill, there would be no deaths during the three months after its application to the population (as was mentioned) and after this time the population would be exposed to the same mortality regime as before the cure. Because for all inhabitants in this model population life was extended for three months, the life expectancy after year *T* will reach the value *70 + 0.25* (i.e. 70 years and 3 months). The situation is completely different during the year *T*, where the number of deaths would be 25 % lower than in the other years. As a consequence of that, the traditionally computed period life expectancy would rapidly increase to the value of nearly 73 years (*ibid*.).

For even clearer picture of the influence of tempo effect we can develop similar example and show in more detail the concrete results. In our model we would use the same assumption – that there only a postponement of deaths caused by the parallel shift of the mortality curve exists but the curve itself would not change.

In our example, we also have a stationary population, for simplifying we suppose, that maximal attainable age is equal to 9 years (or generally 9 time periods). Before any change in mortality the population could be characterized by the life table shown in the Table 18. In this model population, there are 550 persons at the beginning of the year and 100 persons die during that year. The population exposed to the risk of dying is calculated by the usage of the individual data, where we suppose that every person who die during the studied year at age *x* will spend exactly 0.5 years at that age. So the exposures are not taken as the numbers of survivors in the middle of the year as in conventional life tables where individual data hardly could be used. The difference will be significant in further steps of the example and was illustrated already in the Chapter 6. For the construction of the life table we assume the uniform distribution of deaths during the year, where the first person dies just after the beginning of the year and the 10th person dies 1/10 of the year before the end of the studied year. So we use the relation

$$q_x = \frac{2*m_x}{2+m_x},$$

where m_x is the mortality rate at completed age x and q_x is the life table function representing the probability of dying between the exact ages *x* and *x + 1*. The life expectancy (mean age at death) before any mortality change is 5 years for such a population. The calculation of the exposure time was illustrated for the 1st row of this model life table in the Sub-chapter 6.1.1. For the other rows the calculation remains the same.

Table 18: Model example of the tempo effect in a life table, year T – 1

Age	Population (1. 1. of the year)	Exposure time	Deaths	m_x	q_x	p_x	l_x	d_x	L_x	T_x	e_x
0	100	96	10	0.10	0.10	0.90099	1000	99	950	5045	5.04
1	90	86	10	0.12	0.11	0.89011	901	99	851	4094	4.54
2	80	76	10	0.13	0.12	0.87654	802	99	752	3243	4.04
3	70	66	10	0.15	0.14	0.85915	703	99	653	2490	3.54
4	60	56	10	0.18	0.16	0.83607	604	99	554	1837	3.04
5	50	46	10	0.22	0.20	0.80392	505	99	455	1282	2.54
6	40	36	10	0.28	0.24	0.75610	406	99	356	827	2.04
7	30	26	10	0.38	0.32	0.67742	307	99	257	470	1.53
8	20	16	10	0.63	0.48	0.52381	208	99	158	213	1.02
9	10	6	10	1.67	1.00	0.00000	109	109	54	54	0.50

Let’s consider a situation similar to the example of Bongaarts and Feeney (2003; 2006). In our population also the deaths will be postponed from the beginning of the year *T*, so that people will die some time later. The amount of the postponement was taken in that way that 2 deaths which would otherwise occur during year *T* are postponed to the year *T + 1*. So the number of deaths at each age during year *T* is equal to 8. Then 2 deaths are postponed to the next year for each age. What happens to the population from the mortality point of view could be seen in the life table for year *T* (Table 19):

Table 19: Model example of the tempo effect in a life table, year T

Age	Population (1. 1. of the year)	Exposure time	Deaths	m_x	q_x	p_x	l_x	d_x	L_x	T_x	e_x
0	100	97.9	8	0.08	0.08	0.92149	1000	79	976	5875	5.88
1	90	87.9	8	0.09	0.09	0.91295	921	80	897	4899	5.32
2	80	77.9	8	0.10	0.10	0.90232	841	82	817	4001	4.76
3	70	67.9	8	0.12	0.11	0.88873	759	84	734	3185	4.20
4	60	57.9	8	0.14	0.13	0.87076	675	87	648	2451	3.63
5	50	47.9	8	0.17	0.15	0.84586	587	91	560	1803	3.07
6	40	37.9	8	0.21	0.19	0.80907	497	95	468	1242	2.50
7	30	27.9	8	0.29	0.25	0.74922	402	101	372	774	1.92
8	20	17.9	8	0.45	0.37	0.63470	301	110	268	402	1.33
9	10	7.9	8	1.01	1.00	0.00000	191	191	134	134	0.70

The average time which each person (deceased at completed age *x*) would spend alive during the last year of his life was considered to be equal to 0.7 of the year (time period). This average length corresponds to the assumption that at each age two deaths are postponed to the next year. In this situation the life expectancy at birth at year *T* reaches the value 5.88 years. The increase is clearly the consequence of the lower number of deaths during the year *T*.

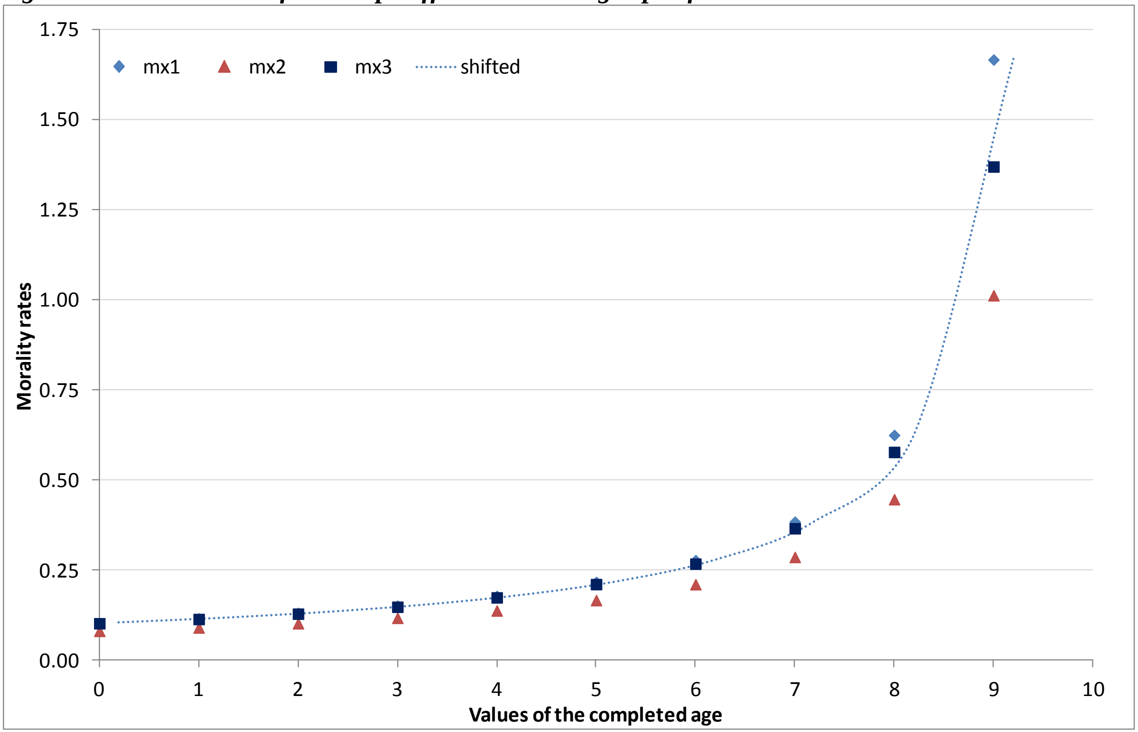
The last sentence from the previous paragraph could be proved by the life table for any year after the assumed mortality change (after year *T*). It could be seen, that in years after year *T* all the deaths remained postponed as happened during year *T*. The average length of life, which was lived by all people deceased at age *x* during that age remained 0.7 year (in comparison to 0.5 before year *T*). In this situation the numbers of deaths at each age are again equal to 10 (2 deaths are postponed to the next year but also 2 deaths are postponed to the studied year from the previous one). Therefore, the numbers of deaths at each age are again the same as before year *T*. What has changed are the exposure times, because thanks to the mortality shift all the deceased persons at age *x* spend 0.2 of the year more alive at that age. The life expectancy decreased again in comparison to the year *T* and without any other supposed change it would remain at this level (see Table 20).

Table 20: Model example of the tempo effect in a life table, year T + 1

Age	Population (1. 1. of the year)	Exposure time	Deaths	m_x	q_x	p_x	l_x	d_x	L_x	T_x	e_x
0	100	97.3	10	0.10	0.10	0.90225	1000	98	971	5301	5.30
1	90	87.3	10	0.11	0.11	0.89166	902	98	873	4330	4.80
2	80	77.3	10	0.13	0.12	0.87849	804	98	775	3458	4.30
3	70	67.3	10	0.15	0.14	0.86169	707	98	677	2682	3.80
4	60	57.3	10	0.17	0.16	0.83949	609	98	580	2005	3.29
5	50	47.3	10	0.21	0.19	0.80880	511	98	482	1425	2.79
6	40	37.3	10	0.27	0.24	0.76359	413	98	384	943	2.28
7	30	27.3	10	0.37	0.31	0.69040	316	98	286	559	1.77
8	20	17.3	10	0.58	0.45	0.55157	218	98	189	273	1.25
9	10	7.3	10	1.37	1.00	0.00000	120	120	84	84	0.70

From the example we can see clearly that the period life expectancy is not the best tool for the analysis of the mortality trends in time. In our model situation mortality as well as the life expectancy were stable before year *T*. Then there was a positive change – postponement of all deaths what would have occurred to older ages. It was corresponding with the parallel shift of the mortality curve to higher ages. The life expectancy reflected that positive development by its increase to the value 5.88. After the year *T* (the year when the mortality change occurs) the situation remained stable again, nothing else happened to the mortality curve. And as a result of this stabilization the value of the life expectancy decreased to 5.30. Many demographers would suppose that there was some worsening of the mortality in the year *T + 1* in comparison to the year *T* when the value of the life expectancy decreased, but it would be a mistake. Mortality conditions and level of mortality did not change at the year *T + 1* or even between year *T* and *T + 1*. The change of the value of life expectancy is only the consequence of the change in timing in the year *T* and the value of life expectancy for that year is overestimated (distorted).

Figure 125: Illustration of the tempo effect on model age-specific death rates



Note: "mx1" are mortality rates at time *T – 1*; "mx2" are mortality rates at time *T*; "mx3" are mortality rates at time *T + 1*; the "shifted" curve represents the real (i.e. defined) change of mortality rates in time *T*

In this moment we can see also the problem of the possible interpretation of the life expectancy mentioned earlier: "the average age at death that would be observed for a group of persons who experience, over the course of their lives, the age-specific death rates observed during the time period" (Bonggarts, Feeney, 2002, p. 14). For the year *T* we can see that the life expectancy is significantly higher in comparison to the previous year and its value corresponds with the observed values of the deaths rates. But as we stated earlier the mortality curve still should be the same after and before the change. It should be only shifted to the right for 0.2 year (0.7 – 0.5). According to the definition, the life expectancy would be the average age at death in the hypothetical cohort where all the mortality rates would be the same as the observed ones during year *T*. The problem is that the observed mortality rates are not the "real" ones (because the "real" ones are defined by our-selves for the model as only shifted ones). From the Figure 125 it could be seen that the mortality curve observed in our example during the year *T* is not the same as before the year *T* shifted only for the 0.20 year to the right.

As a result of these facts it could be stated that some measure of the mean age at death adjusted for tempo effect (as was developed by Bongaarts and Feeney and is described later) should not be taken as an estimation of a cohort measure for any cohort. The adjusted measure remains to be a period measure dealing with some hypothetical cohort (as in the case of traditional period life expectancy) but it should not be distorted by the tempo effects (Barbi *et al.*, 2008; Luy, 2010b). The adjustment methodology and its variants will be presented later in the following text.

Moreover to that, we can show the results when we really suppose one cohort (100 live births) and apply the observed mortality rates from our example to these cohort values. From that data the mean age at death (x_m) could be calculated as

$$x_m = \frac{\sum_{x=0}^{9} x * d_x}{\sum_{x=0}^{9} d_x},$$

where x represents the values of age in our model population and d_x is the distribution of deaths calculated in the form

$$d_x = \frac{D_x}{B},$$

where D_x is the number of deaths at completed age x and B is the number of live births in the model cohort. Because the aim is to apply the estimated age-specific mortality rates (or the estimated probabilities of dying calculated from the rates) to some cohort, we stated the number of births as being equal to 100 and the numbers of deaths (D_x) were calculated using the rates (or probabilities). So as the really calculated rates were used the probability of death for the last age was also calculated from the rate (not set equal to one as in the case of life tables). We suppose no migration in this whole example.

As could be seen for the rates observed in the year T (Table 21), the mean age at death is 5.62, so it is of 0.62 year higher than in the year $T - 1$. But the deaths are shifted only for 0.20 year to higher ages, so the undistorted mean age at death should be 5.20.

Table 21: Model example of the tempo effect in a life table, mean age at death, year T

Age	Population (1. 1. of the year)	Deaths	m_x	q_x	d_x	$x * d_x$
0	100	8	0.08	0.08	0.08	0.0550
1	92	8	0.09	0.09	0.08	0.1364
2	84	8	0.10	0.10	0.08	0.2219
3	76	8	0.12	0.11	0.08	0.3125
4	67	9	0.14	0.13	0.09	0.4098
5	59	9	0.17	0.15	0.09	0.5161
6	50	9	0.21	0.19	0.09	0.6356
7	40	10	0.29	0.25	0.10	0.7763
8	30	11	0.45	0.37	0.11	0.9572
9	19	13	1.01	0.67	0.13	1.2466
					$\sum(x * d_x) / \sum d_x =$	5.6196

When the same calculation is repeated but the observed mortality rates from the year *T + 1* are supposed, the result is completely different. The mean age at death is really equal to 5.20 what the undistorted result is (Table 22).

Table 22: Model example of the tempo effect in a life table, mean age at death, year T + 1

Age	Population (1. 1. of the year)	Deaths	mx	qx	dx	x * dx
0	100	10	0.10	0.10	0.10	0.0684
1	90	10	0.11	0.11	0.10	0.1662
2	80	10	0.13	0.12	0.10	0.2639
3	71	10	0.15	0.14	0.10	0.3617
4	61	10	0.17	0.16	0.10	0.4594
5	51	10	0.21	0.19	0.10	0.5572
6	41	10	0.27	0.24	0.10	0.6549
7	32	10	0.37	0.31	0.10	0.7527
8	22	10	0.58	0.45	0.10	0.8504
9	12	10	1.37	0.81	0.10	0.9482
					$\sum(x * d_x) / \sum d_x =$	5.2000

Another important result could be obtained from our example and it proves the distortion of the values in year *T*. We can calculate the measure which could be called the total mortality rate (TMR; Bongaarts, Feeney, 2002; 2003; 2006). This measure is the corresponding one to the generally known total fertility rate. The total mortality rate could be calculated as the sum of the mortality rates of the 2nd kind (Bongaarts, Feeney; 2006) where there are numbers of deaths in the nominator and numbers of all people ever living in that corresponding hypothetical cohort in the denominator. That means that in the denominator there are numbers of people exposed to the risk of death and also those who already experienced the event. Again the correspondence with the total fertility rate could be hold. The total fertility rate is the sum of the age specific fertility rates where in the nominator are numbers of births and in the denominator are numbers of women in relevant age – women who already have experienced the event of giving birth and also those who have not. In the case of mortality the death rates of the 2nd kind could be calculated as the ratio of the number of deaths at age *x* and the corresponding number of live births. Again no migration was supposed. In our model example we suppose constant numbers of live births each year equal to 100.

Table 23: Model example of the tempo effect in a life table, mortality rates of the 2nd kind, total mortality rate (TMR)

age	year T – 1	year T	year T + 1
0	0.10	0.08	0.10
1	0.10	0.08	0.10
2	0.10	0.08	0.10
3	0.10	0.08	0.10
4	0.10	0.08	0.10
5	0.10	0.08	0.10
6	0.10	0.08	0.10
7	0.10	0.08	0.10
8	0.10	0.08	0.10
9	0.10	0.08	0.10
TMR	1.00	0.80	1.00

Mortality rates of the 2nd kind are equal to 0.10 at each age during the time before and also after the year *T*. The sum of those rates gives the value of one. In the year *T* again the situation is different. The mortality rates of the 2nd kind are equal to 0.08 at each age, so the sum of them is only 0.80 (see Table 23). It could be easily derived what value should be the correct one. The total mortality rate (again in accordance with the total fertility rate) could be interpreted as the total number of events that would experience on average one individual during his life. In case of mortality it is easy to say what value the total mortality rate should equal to. Each person has to (later or earlier) die and each person could die only once during his or her life. So the total mortality rate (with the absence of migration) should be equal to one. Again it could be repeated that this is the proof of the distortion of the traditional period indicators calculated for the year *T*.

9.2 Critical approaches to the concept of tempo effect

As was said already, the concept of the tempo adjustment is a bit controversial in demography especially considering its application to mortality. However, still more and more authors deal with this topic and many methods of extension or modification of the Bongaarts-Feeney method were proposed in theoretical as well as in practical way. The negative reaction of some demographers is not so surprising, because as stated Luy (2010a, p. 410) "such developments are typical for scientific innovations and can be observed in several examples in the history of science" and these critical reactions cannot be taken as a reason for total rejection of this concept (*ibid*.). The defenders of the concept of mortality tempo adjustment suppose that most of the criticism is a consequence of misunderstandings and misinterpretations of the mortality tempo adjustment (Luy, 2010b, p. 416). Luy also sums up four main arguments against the tempo adjustment in the field of fertility (Luy, 2010b, p. 434):

1) "The tempo-adjusted total fertility rate is an inappropriate indicator for cohort fertility,
2) cohort-specific changes in the timing of births are more complex in reality then assumed in the Bongaarts-Feeney formula,
3) the tempo-adjusted total fertility rate does not take into account changes in the parity distributions of the female population,
4) the Bongaarts-Feeney formula is based on inappropriate fertility measures."

In accordance with the previously mentioned main arguments against the tempo-adjustment it could be briefly stated (and was mentioned also in the previous text) that the aim of the tempo-adjusted measure is not to be a cohort indicator. It is common period indicator standardized not only for age but also for tempo.

The second point written above is connected with the assumption of the "constant shape" of the mortality curve. This aspect was mentioned by more scientists (van Imhoff, Keilman, Schoen, Inaba – according to Luy, 2010b). Van Imhoff and Keilman (2000) showed that this assumption does not hold for the Netherlands and Norway. We will deal with this fact also later in this Thesis. As will be shown it is possible to solve it by some extension of the methodology proposed originally by Bongaarts and Feeney (Kohler, Philipov, 2001).

In relation to the third critique it could be stated that this point is not relevant for the mortality process and also already Bongaarts and Feeney (2006) developed another extension of the traditional approach solving this problem (Luy, 2010b). In addition the fourth critique is connected with the process of fertility (*ibid*.).

But the concept of tempo-adjustment is more criticized in case of mortality in comparison to fertility process. Luy (2010b, p. 436) mentioned 2 main critiques related to the tempo-adjustment of mortality measures:

1) "Tempo-adjusted life expectancy is not a suitable measure for cohort life expectancy,
2) mortality tempo adjustment is paradox since life expectancy is a tempo measure itself and the quantum of mortality is always one by definition."

The first argument was mentioned e.g. by Goldstein (2006). He concludes that "[w]hen mortality conditions are improving period life expectancy is less than that of the cohort born in the period. This is because the hypothetical cohort following the period life table is deprived of future mortality improvement." (Goldstein 2006, p. 72). His results are opposite to the results of Bongaarts and Feeney (2002). However, he himself argues that Bongaarts and Feeney did not deal with cohorts and that they take the tempo-adjusted measure as a period one. Goldstein (2006) also proved that under the assumption of linear shift[16] the tempo-adjusted life expectancy is equal to the life expectancy of the cohort dying in the studied year. That also clarifies the results of Bongaarts and Feeney who expected the tempo-adjusted life expectancy to be less than the current unadjusted one. Moreover, this critique could be explained by the same argument as in the case of fertility, i.e. that it is not the aim of the tempo-adjusted measures to supplement the cohort ones.

The second critique was mentioned e.g. by Wachter (2005). He argues (Wachter, 2005, p. 202), that "[i]n the study of mortality, no distinction between quantum and tempo exists at the individual level. A person has one death, his or her own, and mortality pertains to whether death comes early or late. It makes no obvious sense to adjust away the effects of changes in the timing of death, thus adjusting away changes in mortality itself". He also mentions that the tempo-adjusted indicator measures rather earlier mortality conditions while the current mortality conditions are measured by the standard period life expectancy. Luy (2010b) argues that the tempo effects are the result of the empirical age-specific mortality rates which are tempo-distorted and not of the life tables as was supposed by the critiques. It was also visible in the model examples above that the distortion affects already the age-specific mortality rates (Figure 125).

Rodrígues (2008) argues that the proposed method of adjustment has not similar importance and sense for the process of fertility as it has for mortality. In fertility it could, according to Rodrígues, enable to distinguish between the quantum (real intensity of the process) and tempo (in the meaning of changes in timing of the events during the life cycle). While in the mortality process, there the quantum is fixed (every person has to experience death only once in his or her life) and only the tempo could change (as people can die later or earlier). So in mortality analysis it is not possible to mistake tempo for quantum and vice versa (*ibid*.).

On the other hand, Rodrígues deals with the interpretation of the adjusted measure. He pointed out that the tempo-adjusted measure combines the observed force of mortality function and also the age structure of the population, what is the consequence of the past development and not of

---

[16] The assumption of the linear shift supposes that the deaths are not only postponed uniformly across ages but also that the postponements (shifts) are constant in all years (Goldstein, 2006).

the current mortality conditions. He then concludes, that the difference between the traditional period life expectancy and the adjusted one is not the result of some tempo distortion, he supposes that both these indicators simply measure different things – the adjusted measures (for the process of fertility) "tell us how many children the synthetic cohort would have (and when) if it followed a shifting period schedule with constant shape and quantum but changing tempo" (Rodrígues, 2008, p. 76).

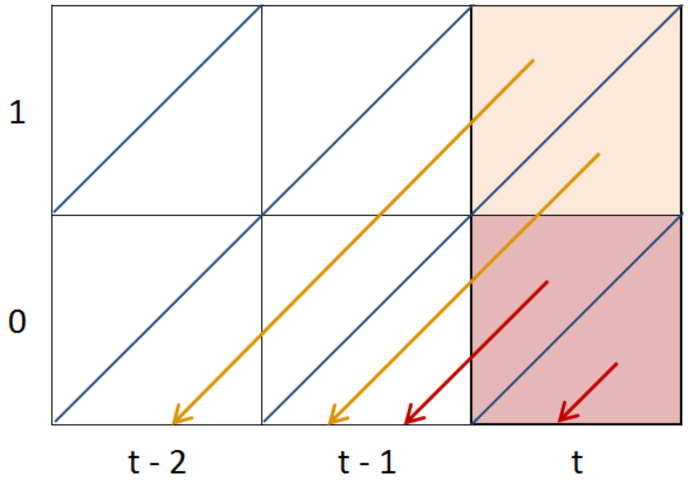
There could be mentioned many other critics of the tempo-adjusted measures as well as its defenders. In the text below the concept will be studied without any clearly defined conviction. The most important assumptions on which the tempo-adjustment is based are verified and more different methods will be compared.

9.3 Summary of the main equations used in the analysis and an analysis of the real data

Although some basic formulas used within this part of the work were already presented in the previous chapter, in this place there will be partly repeated and the concrete way of calculation used for the real mortality data will be described. It is not possible to describe the tempo effects and its consequences only through a theoretical example, so similar example as in the previous chapter will be used here but the input data were taken from the Human Mortality Database (Human Mortality Database, 2010). For the first example we use mortality data for Swedish females during the years 1987 to 2007 when the deaths postponement could be assumed. We have to start by calculating the deaths rates of the 2nd kind (reduced death rates). As was defined earlier, we suppose the death rate of the second kind to be the ratio of the number of deaths at age *x* during year t and the corresponding numbers of live births (born in year *t* – *x*):

$$d^r_{(x,t)} = \frac{D_{(x,t)}}{B_{(t-x)}},$$

Figure 126: Lexis diagram illustrating the calculation of the reduced mortality rates



where the symbol $d^r_{(x,t)}$ denotes the reduced mortality rate, $D_{(x,t)}$ is the number of deaths that occur at age *x* during the year *t*, and $B_{(t-x)}$ is the corresponding number of live births. There the question arises what is the "corresponding number of live births"? Because the 3rd classes of deaths are used (as showed in the Lexis diagram in the Figure 126) the deceased persons in each rectangle were originally members of two different cohorts. For the age 0 we can clearly see that persons who died at age 0 during year *t* were born during year *t* or *t – 1*. Situation is the same for all ages. To handle this problem we can rewrite the formula for the reduced mortality death rate to the form:

$$d^r_{(x,t)} = \frac{D_{(x,t)}}{(B_{(t-x)} + B_{(t-x-1)})/2}.$$

In accordance to Bongaarts and Feeney and to the aim of this Thesis we omit the mortality at ages lower than 30 and concentrate only on the adult mortality at middle and higher ages.

Because in the Human Mortality Database numbers of deaths are published up to age 110 we need quite a long time series of births if we want to use all the deaths. This is not a problem for countries like Sweden where long time series is available.

At first we have to return to the theoretical roots of the model as were defined by Bongaarts and Feeney (1998; 2002; 2003; 2006). They suppose (and it was emphasized above within the model example) the parallel shift of the mortality curve to higher ages as the only change in mortality that happened. This assumption was easy to fulfill in a theoretical model, but in real data (as was shown in the Chapter 8) the pure and undistorted shift only rarely happens. It was also mentioned by Luy (2010b). Usually the mortality change is not proportional according to age. In the article of Yi and Land (2001) it was proved that the method proposed by Bongaarts and Feeney is robust enough for being used also in the case of slightly changing variability of the hazard curve. The same was proved also by Kohler and Philipov (2001) what will be mentioned later in more detail.

We start our analysis for Sweden by computing the mortality rates of the second kind according to equation introduced earlier. It must be pointed out that Bongaarts and Feeney (2006) define the mortality rate of the 2nd kind as the numbers of deaths at age *a* and time *t* divided by the number of all persons born at time *t – a* (Bongaarts, Feeney, 2006, pp. 118; Bongaarts, 2012). However, they obtained the deaths by age after age 30 by calculating a hypothetical period and cohort life tables in which mortality was set to zero below age 30 (Bongaarts, 2012). We will use the real data in our analysis. Therefore we have to subtract the numbers of deaths before age 30 from the numbers of births in the denominator. Subtracting the deaths at ages below 30 from the numbers of births is the same as to use the numbers of survivors at the age of 30. If we did not do that, the sum of the rates of the 2nd kind (the total mortality rate) could never reach the value of one because the deceased persons at age lower than 30 would be included in the number of births in the denominators but they would be omitted in the nominators. In accordance to that, within this chapter the mortality rates of the 2nd kind will be calculated as stated below (where the deaths at ages below 30 are subtracted in the nominator as well as in the denominator). That means that in the denominator the numbers of survivors at age 30 from each cohort are taken into the calculation (migration is neglected because no data about migration were available):

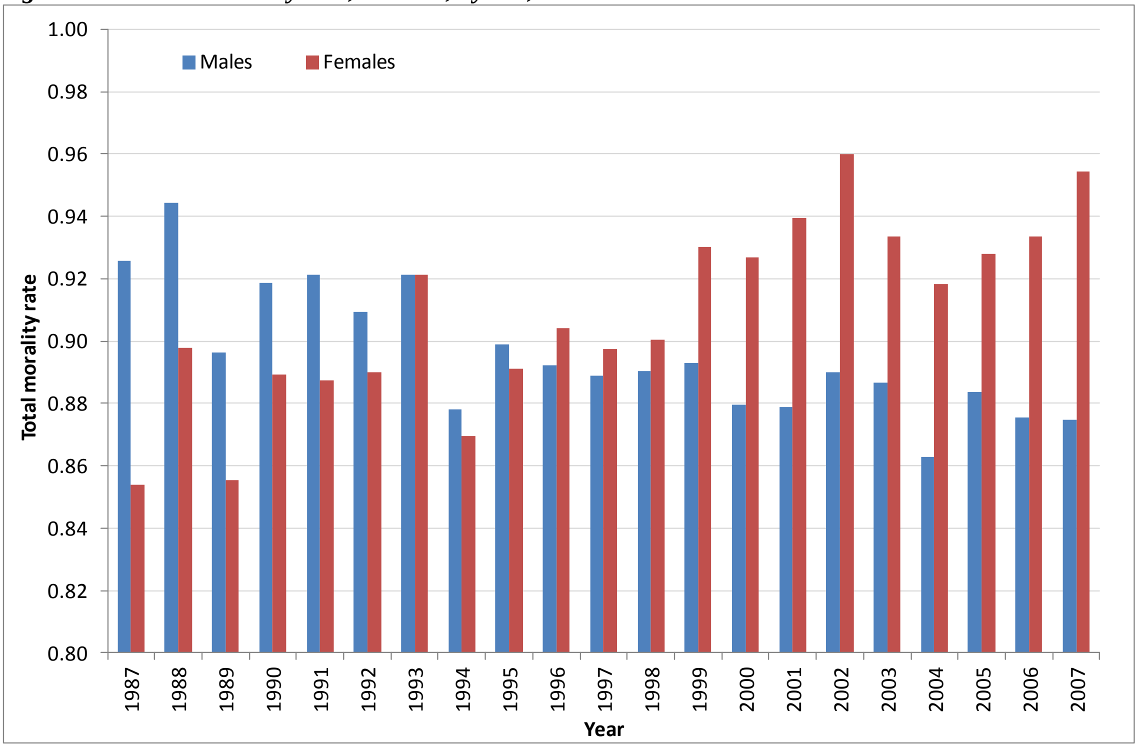
$$d^r_{(x,t)} = \frac{D_{(x,t)}}{(P_{30}^{t-x} + P_{30}^{t-x-1})/2}$$

for all ages x equal at least 30, where P_{30}^{t-x} is the number of survivors to age 30 in a cohort born in the year $t - x$. The sum of the mortality rates is the total mortality rate:

$$TMR_t = \sum_{x=30}^{\omega} d_{(x,t)}^{r}.$$

It was confirmed that the total mortality rate is less than one as a consequence of increasing mean age at death during the studied years. Although for females the value of total mortality rate is on average increasing, during the latest years almost stabilization is visible. As was derived already by Ryder (1956) the difference between the calculated total mortality rate and the expected value of this measure (i.e. one) could be taken as the amount of the distortion caused by the tempo effect. From the Figure 127 it could be supposed that the amount of tempo distortion is decreasing for females and reaches relatively low values already. This result corresponds with the conclusions about the occurrence of the almost pure mortality shifting for Swedish females during the last years (see Chapter 8, Figure 94). However, some tempo distortion still remains there.

Figure 127: Total mortality rate, Sweden, by sex, 1987–2007



Source of data: author's calculation based on Human Mortality Database (2010)

In comparison to fertility, in the case of mortality we are not interested in the tempo-adjusted total event rate (total mortality rate or total fertility rate) because this adjusted rate should be equal to one. What we are interested in it is the tempo-adjusted mean age at the event that means the tempo-adjusted mean age at death. As was shown in Bongaarts and Feeney (2006) the mean age at death based on the rates of the 2nd kind is not distorted and so it needs no adjustment.

In practice the rates of the 2nd kind are not usually used for the estimation of the mean age. Traditionally the mean age at death is estimated by the life expectancy at birth calculated in period life tables. But it was shown already that the period life expectancy is distorted by changes in tempo (timing) of mortality. In the next step we will compute the mean age at death based on the rates of the 2nd kind (according to Bongaarts and Feeney, 2006) and we can compare the results with the traditional period life expectancy. The mean age at death based on the rates of the 2nd kind will be labeled as *MAD* – the mean age at death:

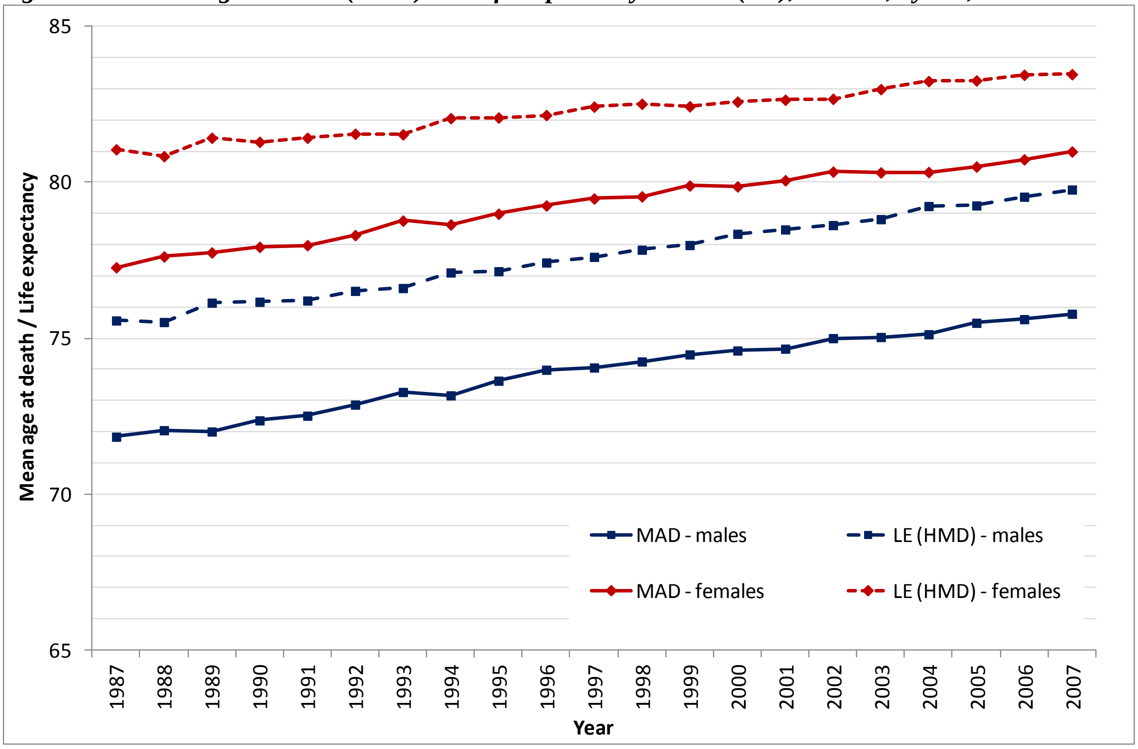
$$MAD(t) = \frac{1}{TMR(t)} * \int_0^{\infty} x * d^{r}_{(x,t)}\, dx$$

In the Figure 128 the results could be seen clearly. The life expectancy at birth originates from the Human Mortality Database (labeled as "*LE (HMD)*"). In accordance with the assumption that there are no deaths below age 30, the life expectancy at birth was calculated as:

$$LE(HMD) = e_{30} + 30.$$

In both cases, for males as for females, the life expectancy is higher significantly than the mean age at death calculated on the basis of the rates of the 2nd kind (labeled as "*MAD*"). The difference is around 3 years and it is decreasing in the case of females what contributes to the results of decreasing distortion of total mortality rate, presented earlier.

Figure 128: Mean age at death (MAD) and life expectancy at birth (LE), Sweden, by sex, 1987–2007



Source of data: author's calculation based on Human Mortality Database (2010)

As Bongaarts and Feeney (2006) pointed out, the rates of the 2nd kind could be affected by compositional effects. Therefore, they propose the usage of rates of the 1st kind and a measure of the mean age derived from this type of rates. However, it is important to keep in mind that these rates are affected by the tempo effects and therefore need the adjustment.

Rates of the 1st kind are frequently used while constructing the life table. That means that the mean age at death based on the rates of the 1st kind is the standard traditionally computed period life expectancy, we will label it in accordance to Bongaarts and Feeney (2006) as $MAD_L(t)$,

$$MAD_L(t) = e_0(t) = \int_0^{\infty} exp\left[-\int_0^{a} \mu(x,t)dx\right]da,$$

where t denotes time, x is the age and $\mu(x,t)$ is the hazard rate at age x and time t. This type of measure needs, however, adjustment because the rising mean age depresses rates of the 1st kind and that is why also the life expectancy (Bongaarts, Feeney, 2006). Bongaarts and Feeney (2006) stated that the means of the 1st kind are higher than those of the 2nd kind. The difference equals the tempo effect.

The proposed way of adjusting of the measure based on the rates of the 1st kind holds only under the "constant shape assumption". Bongaarts and Feeney understood the constant shape assumption as being equal to the parallel shift of the survival function to higher or lower ages. Then the distortion could be removed from the measures of the 1st kind simply by dividing the numerators of the hazard rates from which they are derived by $\left(1 - r_p(t)\right)$, where $r_p(t)$ denotes the rate of change in the period mean age at death. The authors proposed to use the rate of change in the mean age of the second kind to calculate the distortion index needed for the adjustment. So the formula for the tempo adjusted mean age at death based on the rates of the 1st kind could be derived as (Bongaarts, Feeney, 2006):

$$MAD_L^*(t) = e_0^*(t) = \int_0^{\infty} exp\left\{-\int_0^{a} \frac{\mu(x,t)}{TMR(t)}dx\right\}da,$$

where *TMR (t)* is the total mortality rate computed and described in the previous part. Again we suppose that the intensity of mortality below age 30 is equal to zero.

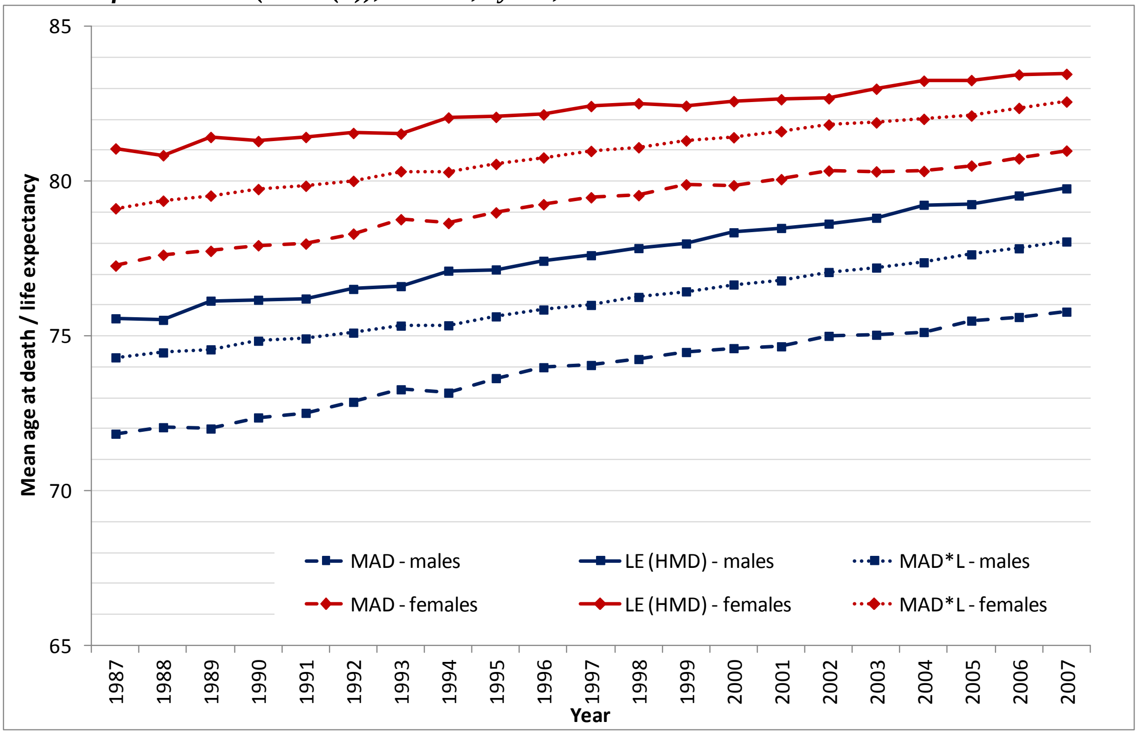
From the results of Bongaarts and Feeney (2006, p. 16) one could see that the mean age at death based on the rates of the 1st kind lies close to the mean age based on the rates of the 2nd kind for the selected countries. This closeness was proved under the constant shape assumption (Bongaarts, Feeney, 2003). The authors concluded also that under this assumption it holds even:

$$MAD_L^*(t) = MAD(t).$$

According to our example on Swedish data shown in the Figure 129 there are visible differences of our results from the assumed ones. The differences could be taken as the consequences of not exactly parallel shift of the survival curve in the real population and also the consequence of a slightly different calculation of the mortality rates of the 2nd kind (explained above).

It could be concluded that in our model example the assumed relations were verified and the tempo distortion of the period life expectancy could be taken as to be proved. However, it was shown that for the real example of Sweden at the end of the 20th century and at the beginning of the 21st century for both sexes the results are partly different from the results of the original authors of the concept. In the next part we will consider also more historical data where the constant shape assumption could be taken as even less realistic in comparison to the contemporary Swedish situation. Although it was proven by Yi and Land (2001) and Kohler and Philipov (2001) that the proposed method of adjustment is robust enough for being used also when the shape of the mortality curve is not strictly constant, we can suppose that for more historical data the shape of the mortality and survival curves could have changed more significantly and that it could more significantly influence the results.

Figure 129: Mean age at death (MAD), life expectancy at birth (LE) and mean age at death based on the rates of the 1st kind (MAD (L)), Sweden, by sex, 1987–2007



Source of data: author's calculation based on Human Mortality Database (2010)

9.4 Application of the Bongaarts-Feeney adjustment to historical data

The aim of this part is to verify the results produced by the Bongaarts-Feeney method of tempo adjustment (mentioned above) for as historical data as possible where we can rightfully assume that the constant shape assumption will not hold anymore. Yi and Land (2001) and Kohler and Philipov (2001) supposed the Bongaarts-Feeney method to be robust enough. However, in these articles mostly the contemporary data were used for its testing.

Data from the Human Mortality Database were used in this section. Only for several countries it was possible to calculate the adjusted measures. The reason is the data-demands of the method. While we are using mortality data up to age 110, we need also the numbers of live births for at least 110 years backwards. That is why the analysis was realized for 5 countries only[17] where the time series is long enough not only for the calculation of tempo-adjusted measures but also for

[17] Those are Sweden, France (civilian population), Denmark, Belgium, and Norway.

more illustrative presentation of the results. Firstly, for all those countries the mortality rates of the 2nd kind were calculated and the total mortality rate as the sum of these rates.

$$d^r_{(x,t)} = \frac{D_{(x,t)}}{(P_{30}^{t-x} + P_{30}^{t-x-1})/2},$$

where P_{30}^{t-x} is the number of survivors to age 30 in a cohort born in the year $t - x$.

$$TMR_t = \sum_{x=30}^{\omega} d^r_{(x,t)}.$$

Again the numbers of deaths and the values of the mortality rates were assumed to be zero for ages below 30. Then the mean age at death based on the rates of the 2nd kind will be calculated for all the countries in the analysis:

$$MAD(t) = \frac{1}{TMR(t)} * \int_0^{\infty} x * d^r_{(x,t)}\, dx.$$

The adjusted mean age at death based on the rates of the 1st kind could be estimated through the formula:

$$MAD_L^*(t) = e_0^*(t) = \int_0^{\infty} exp\left\{- \int_0^{a} \frac{\mu(x,t)}{TMR(t)} dx\right\} da,$$

where we assume that $\mu(x + 1/2, t) \cong m(x,t)$.

Finally for comparison we will use the life expectancy from the Human Mortality Database, $LE(HMD)$. Because we assume the mortality under 30 to be equal to zero, we have to use the life expectancy calculated as:

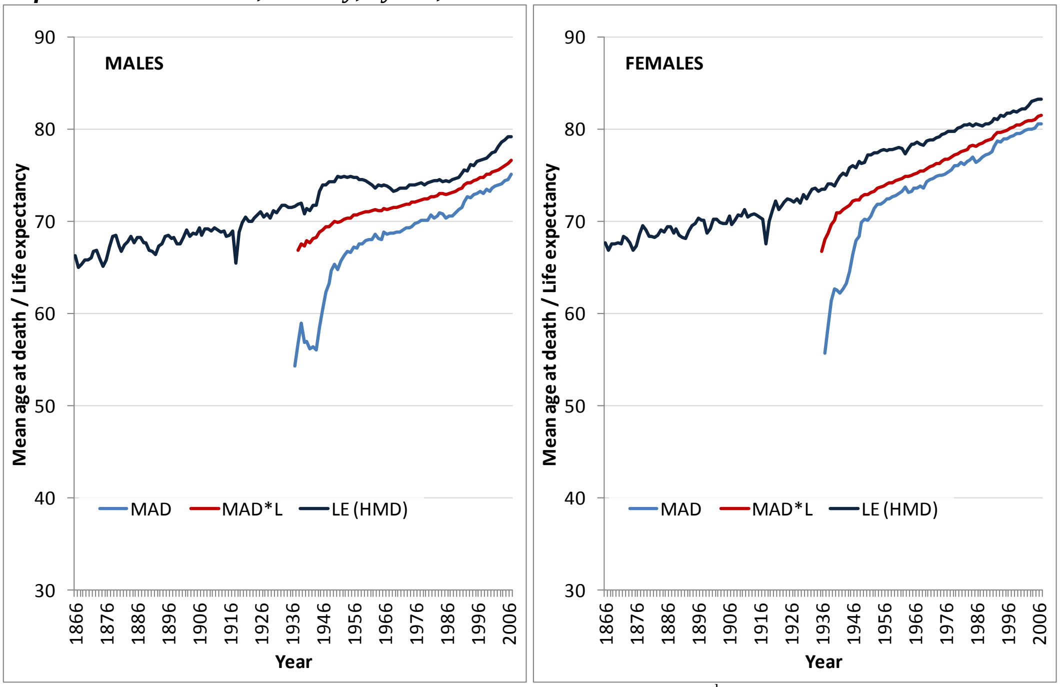
$$LE(HMD) = e_{30}(t) + 30.$$

According to the Bongaarts-Feeney methodology applied to the data the results should have respect the assumptions mentioned above. But it was shown that the relation

$$MAD_L^*(t) = MAD(t)$$

does not hold for the studied populations and above all for more historical data. We can assume that this result is the natural consequence of the fact that the constant shape assumption does not hold for the less developed (historical) populations. The other reason could be again found in the slightly different method of calculation (see the previous sub-chapter). In almost all cases it holds true that the traditional period life expectancy remains higher than the mean age at death based on the rates of the 2nd kind, $MAD(t)$. The adjusted mean age at death based on the rates of the 1st kind, $MAD_L^*(t)$, is significantly different from the $MAD(t)$. In the latest decades those two curves are approaching each other. The reason for that trend could be the tendency of mortality changes towards the mortality shifting defined by the parallel move of the survival curve.

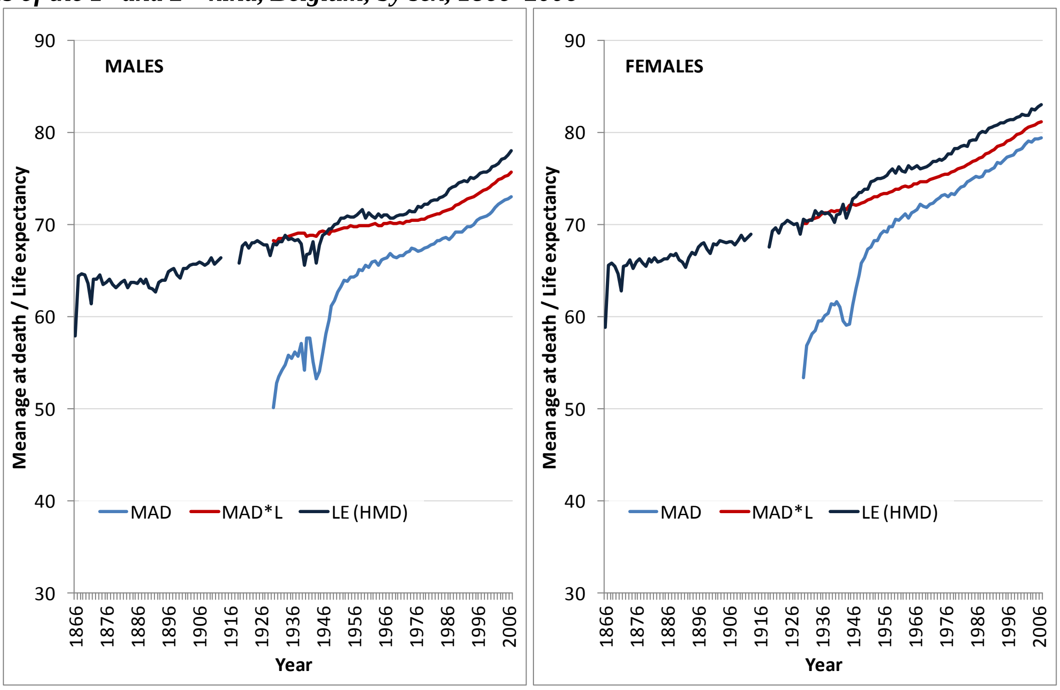
Figure 130: Life expectancy at birth compared with the tempo-adjusted mean age at death based on the rates of the 1st and 2nd kind, Norway, by sex, 1866–2006



Note: "MAD" is the adjusted mean age at death based on the rates of the 2nd kind; "MAD*L" is the adjusted mean age at death based on the rates of the 1st kind; "LE(HMD)" represents the life expectancy at birth (calculated as the life expectancy at age 30 increased by 30) from the Human Mortality Database

Source of data: author's calculation based on Human Mortality Database (2010)

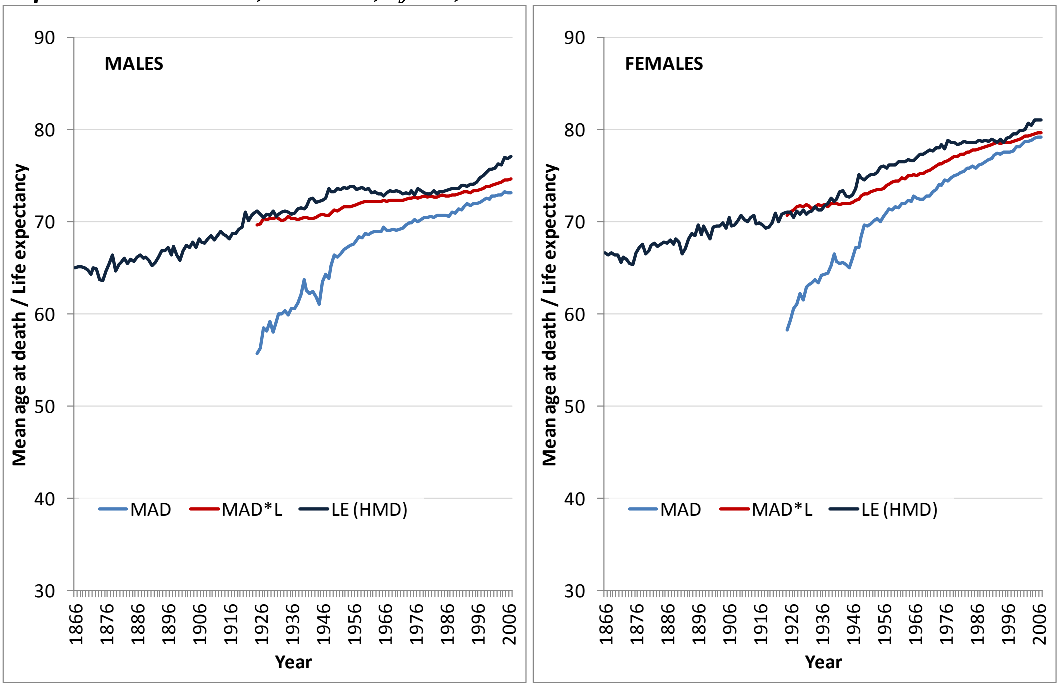
Figure 131: Life expectancy at birth compared with the tempo-adjusted mean age at death based on the rates of the 1st and 2nd kind, Belgium, by sex, 1866–2006



Note: see Figure 130

Source of data: author's calculation based on Human Mortality Database (2010)

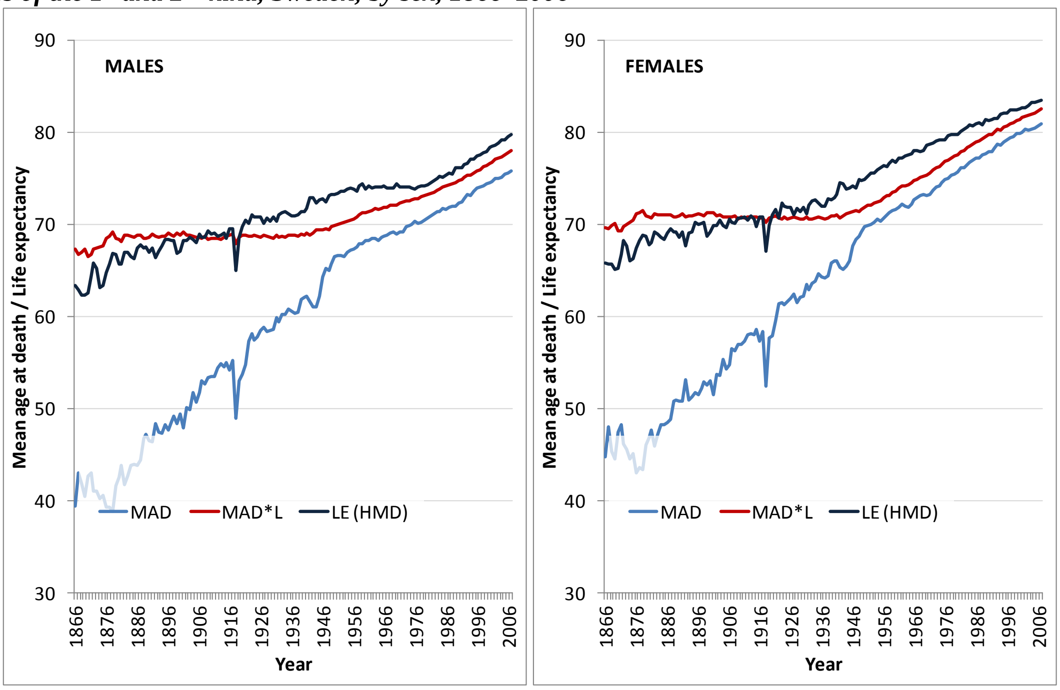
Figure 132: Life expectancy at birth compared with the tempo-adjusted mean age at death based on the rates of the 1st and 2nd kind, Denmark, by sex, 1866–2006



Note: see Figure 130

Source of data: author's calculation based on Human Mortality Database (2010)

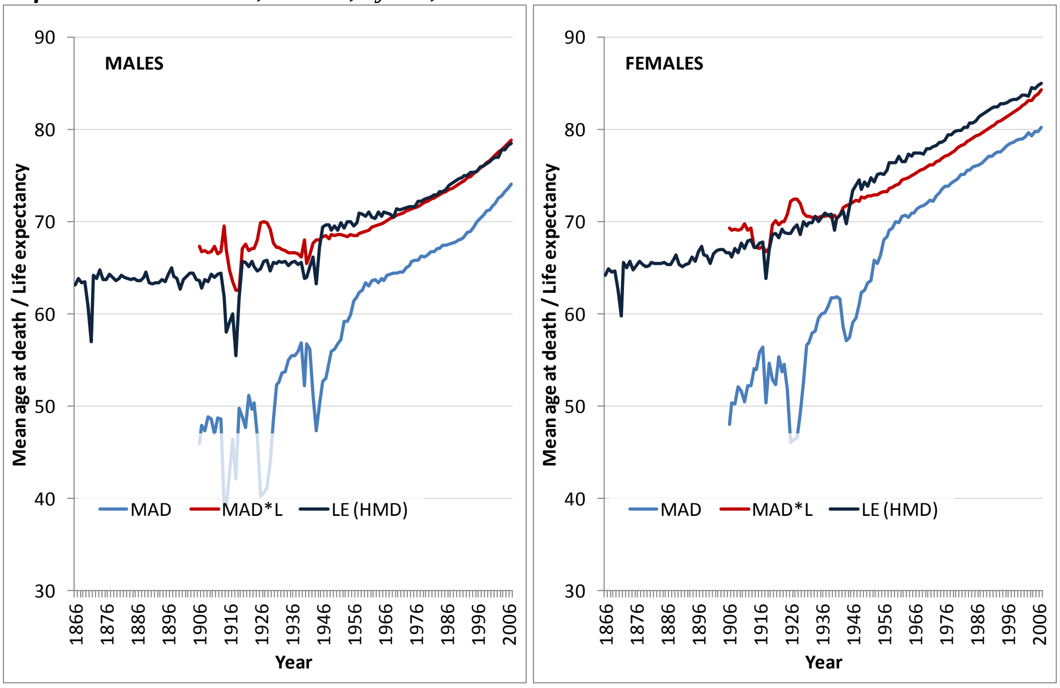
Figure 133: Life expectancy at birth compared with the tempo-adjusted mean age at death based on the rates of the 1st and 2nd kind, Sweden, by sex, 1866–2006



Note: see Figure 130

Source of data: author's calculation based on Human Mortality Database (2010)

Figure 134: Life expectancy at birth compared with the tempo-adjusted mean age at death based on the rates of the 1st and 2nd kind, France, by sex, 1866–2006



Note: see Figure 130

Source of data: author's calculation based on Human Mortality Database (2010)

9.5 Adjustment of the Bongaarts-Feeney methodology for changes in variability of mortality

As was stated above the Bongaarts-Feeney methodology is based on the constant shape assumption. However, (especially in history or in developing countries) this assumption could not be taken as real. In such populations the variability of the distribution of deaths changes significantly (as was shown in Chapters 7 and 8).

Therefore, there arises a question whether there is another possible modification of the proposed methodology which could handle the changes of variability in mortality (changes in the shape of the mortality curve). Such an attempt was made by Kohler and Philipov (2001). They did not deal with the tempo adjustment in mortality but their paper was a response to Bongaarts and Feeney (1998) and to their proposal of tempo adjustment of the total fertility rate. Kohler and Philipov (2001) presented an extension of the Bongaarts-Feeney's methodology which included also the variance effects (changes of variance of the fertility schedules). They supposed that the increase in the mean age at birth is frequently accompanied by an increase in the variance of the fertility schedule (*ibid*.). Also in mortality we can reasonably assume some changes of the variability while the mean age at death is rising. So we can suppose that some adjusted methodology would be useful also in the case of mortality.

The aim of this part is to briefly introduce the modified methodology proposed by Kohler and Philipov (2001). Then this methodology will be applied to our data (introduced and analyzed in the previous part). There are theoretically 3 possible results of the comparison of our previous results with this last modification:

- First, the results obtained by the Kohler-Philipov methodology will be similar to the measure based on the rates of the 1st kind, $MAD_L^*(t)$. In that case it would mean that the $MAD_L^*(t)$ is not distorted by variance and the measure based on the rates of the 2nd kind is under- or over-estimated because the changes of variance of the mortality schedule.
- The second possible result can show that the results of the Kohler-Philipov methodology would be close to the mean age at death based on the rates of the 2nd kind. It would mean that the measure based on the rates of the 2nd kind is relatively robust to the changes in mortality schedules while the $MAD_L^*(t)$ is not.
- The last possible result would be in the case that the variability- and tempo-adjusted mean age at death (according to the Kohler-Philipov methodology) would be different from both previously mentioned measured ($MAD_L^*(t)$ and also $MAD(t)$). In that case it would be possible to assume that both previously mentioned measures are distorted by the changes in the mortality schedules – supposing that the Kohler-Philipov formula is not distorted and is correct.

Based on the results of the following analysis we will finally summarize the assumed facts about tempo effects and variance effects and the adjusted measures will be presented together with the interpretation of the results.

9.5.1 Kohler-Philipov methodology

Kohler and Philipov (2001) developed a general framework of fertility postponement. In this framework the changes in the tempo of fertility can depend on both – period and age. They derived, above several assumptions and partial results, a procedure that iteratively corrects the estimation of the mean age at death for the effect of the changing variance. Their methodology will be described briefly and used in the following part.

Before the beginning of the first iteration four parameters should be estimated. One of them is the observed mean age at the studied event (labeled unusually as $\mu(t)$) – in the original paper this is the mean age at the first birth. When we transfer the procedure to the process of mortality, this first estimate of the observed mean age at death is equal to the above defined $MAD(t)$. The second parameter which has to be estimated is $\sigma^2(t)$, the observed variance, which could be calculated as

$$\sigma^2(t) = \frac{\sum_{x=0}^{\infty}\left[d_{(x,t)}^{r} * (x - MAD(t))^2\right]}{TMR(t)}.$$

The third parameter needed in the procedure is the third centralized moment of the fertility schedule (in the original paper) in year t, labeled as $\kappa(t)$. According to the principles of the statistical moments, this could be calculated as

$$\kappa(t) = \frac{\sum_{x=0}^{\infty}\left[d_{(x,t)}^{r} * (x - MAD(t))^3\right]}{TMR(t)}.$$

The last estimate needed before the first iteration is the estimate of the rate of change in the variance, $\hat{\delta}(t)$. Kohler and Philipov (2001) proposed a simple formula for this estimation in the form:

$$\hat{\delta}(t) = 0.25 * log\left(\frac{\sigma^2(t+1)}{\sigma^2(t-1)}\right).$$

Kohler and Philipov (2001) also recommended using the smoothed time series of the variance and mean age at death. That means that both these time series should be smoothed before we start the iteration process. In the later calculation this requirement showed to be very important otherwise the results are very unstable and sometimes also incalculable.

In the first step of the Kohler-Philipov estimation process the first estimate of $\hat{\gamma}_0(t)$, the change of the mean age at the event, could be calculated as

$$\hat{\gamma}_0(t) = \frac{1}{2} * [\mu(t+1) - \mu(t-1)]$$

and the variance of the fertility schedule (originally; in our case it is the mortality schedule) as

$$\hat{s}_0^2(t) = \sigma^2(t).$$

Just after this preparation phase the main iterative procedure begins. This procedure consists of three connected steps:

1) the corrected estimation of the variance of the schedule

$$\hat{s}_n^2(t) = \sigma^2(t) + \left[\frac{\hat{\delta}(t)}{1-\hat{\gamma}_{n-1}(t)} * \hat{s}_{n-1}^2(t)\right]^2 + \frac{\hat{\delta}(t)}{1-\hat{\gamma}_{n-1}(t)} * \kappa(t),$$

where *n* signifies the number of the iteration;

2) the corrected adjusted mean age at the event

$$\hat{a}_n = \mu(t) + \frac{\delta(t)}{1-\hat{\gamma}_{n-1}(t)} * \hat{s}_n^2(t);$$

3) the corrected estimate of the tempo effect

$$\hat{\gamma}_n(t) = \frac{1}{2} * [\hat{a}_n(\mathrm{t}+1) - \hat{a}_n(\mathrm{t}-1)].$$

These three steps are repeated until the estimates converge. The final step of the procedure then means only the calculation of the adjusted total fertility rate with variance effects. Kohler and Philipov (2001) proposed the calculation as $\frac{TFR(t)}{1-\hat{\gamma}(t)}$.

9.5.2 Application of the Kohler-Philipov methodology

For the purpose defined above the Kohler-Philipov methodology described in the previous sub-chapter will be applied to the empirical data used already in the previous part of the chapter. During the application it showed that the time series of the mean age at death ($MAD(t)$) and variance ($\sigma^2(t)$) have to be smoothed as Kohler and Philipov (2001) proposed. In our analysis the smoothed average of the degree of five was used.

Moreover to that it revealed that an additional correction is needed. In the cases were deep and rapid changes in the values occurred these variations had to be replaced by the smoothed values (the interpolation of these values was made by the weighted averages of the values before and after the deviation). After this correction the values were smoothed by the moving average and only then the described procedure was applied.

So that the results could be compared, all the data for the above selected 5 countries (as in the previous calculation) were used. Again the input data come from the Human Mortality Database.

Kohler and Philipov (2001) proposed also another important fact for the calculation. It is the fact preventing shortening of the resulting time series. Because in each iteration the time series becomes shorter for the first and the last value the authors of the methodology proposed to add the first and the last value in each iteration as equal to the second value and to the last but one value. We followed all those proposals in the application of this model within this work.

9.5.3 Modification of Kohler-Philipov methodology for the needs of mortality

As was described the Kohler-Philipov methodology in the previous section, it was clear that this method was developed for the purpose of the adjustment of the total fertility rate and the mean age at the first birth. The difference which could be used in the case of mortality in comparison to fertility is the fact that we know what the value of the adjusted total mortality rate should be. This value should be equal to one, because each person has to die one time and each person could die only once. The fact that the adjusted total mortality rate (TMR_{adj_var}) could be taken as equal to one could be used for the simplification of the described computational process.

This simplification could be proposed in several steps:

If we suppose that $TMR_{adj\_var}(t) = 1$, then it could be written

$$TMR_{adj\_var}(t) = \frac{TMR(t)}{1-\hat{\gamma}(t)}.$$

If these facts are accepted the estimation consists of the following steps:

1) the estimation of the mean age at death ($MAD(t)$), the rate of change of variance, $\hat{\delta}(t)$, the third centralized moment of the mortality schedule (in the original paper it was the fertility schedule) in year *t*, $\kappa(t)$, and the rate of change of variance, $\hat{\delta}(t)$. All the formulas needed for these estimates remain the same as were described above;
2) the change of the mean age at death (the tempo effect) then could be estimated simply as

$\hat{\gamma}(t) = 1 - TMR$(t) and

the estimation of the variance of the mortality schedule:

$\hat{s}_0^2(t) = \sigma^2(t);$

3) the corrected estimation of the variance should be equal to

$$\hat{s}^2(t) = \sigma^2(t) + \left[\frac{\hat{\delta}(t)}{1-\hat{\gamma}(t)} * \hat{s}_{n-1}^2(t)\right]^2 + \frac{\hat{\delta}(t)}{1-\hat{\gamma}(t)} * \kappa(t);$$

4) then finally the estimate of the adjusted mean age at death could be assumed to be equal to

$$\hat{a}\,(t) = MAD\,(t) + \frac{\hat{\delta}(t)}{1-\hat{\gamma}(t)} * \hat{s}_n^2(t).$$

This slightly modified and simplified methodology of Kohler and Philipov was applied to the empirical data from all the selected countries. The results will be compared with the previous ones with the aim to reveal the importance of the variance-distortions of the tempo-adjusted indicators proposed by Bongaarts and Feeney.

9.5.4 Results and the brief summary of Kohler-Philipov methodology

From the results shown in the Figures 135–139 it is clear that one of the supposed result really occurred – the mean age at death ($MAD(t)$) lies for all the analyzed countries very close to mean age at death adjusted for variance (calculated by the Kohler-Philipov methodology and labeled as "*MAD (adj. variance) n = 20*"[18]) and also close to the simplified Kohler-Philipov methodology proposed in the previous sub-chapter of this work for the process of mortality (labeled as "*MAD (adj. variance) – simplified*"). The only measure which differs from those three is the mean age at death based on the rates of the 1st kind ("*MAD*L*"). Therefore, it could be supposed that the mean age at death based on the rates of the 2nd kind is relatively robust to the changes in variance, because the results are similar to the values of measures adjusted for the variance.

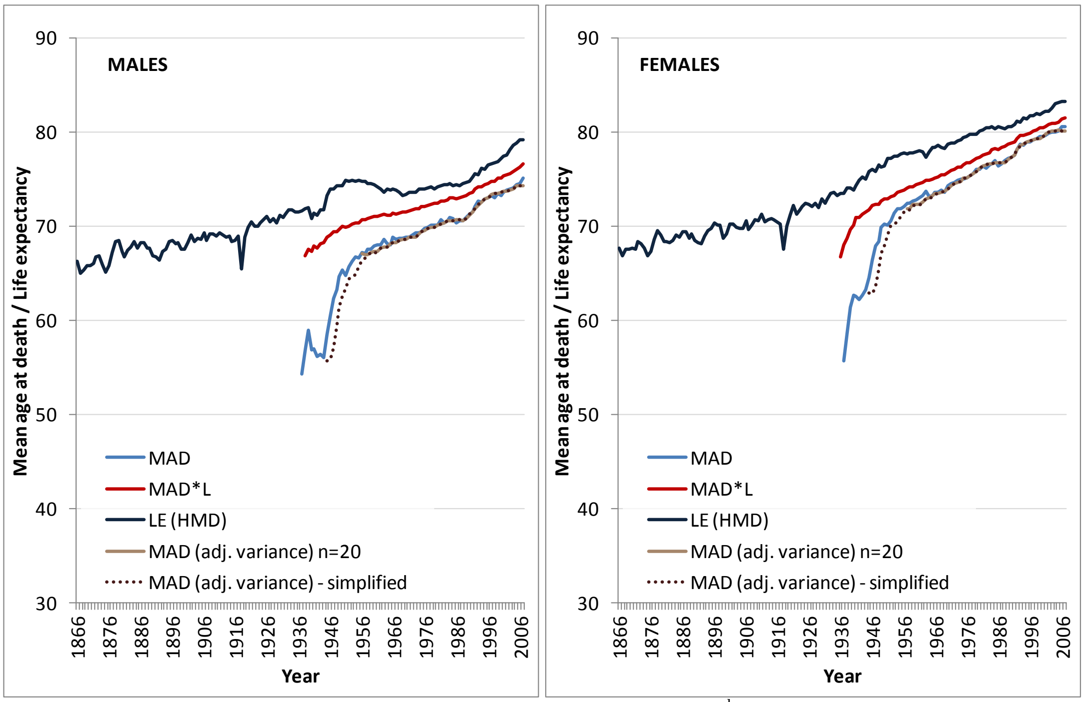
On the other hand it seems that the measures based on the rates of the 1st kind are more distorted by the variance and its changes. That is the reason why all those measures are approaching during the last years in modern and demographically developed countries, because (as was showed in the chapter devoted to the mortality shifting) the change of mortality in those developed countries approaches the parallel shift of the survival curve as was the initial assumption made by Bongaarts and Feeney (1998; 2002; 2003 and 2006).

From the results shown in all the following figures it is clear that the tempo-adjusted mean age at death (*MAD*) was in all the cases lower than the traditionally used life expectancy. It is in

[18] The „*n = 20*“ signifies that 20 iterations were done in the calculation.

accordance with the initial assumption that when the mean age is increasing there is a tendency to overestimation of the life expectancy. In the past also a higher variability could be seen (it is visible e.g. for the population of Sweden). During the past decades the development is smoother and all the curves are more similar. As the improvement of the mortality slows down the traditional life expectancy is probably less distorted by the tempo. In the same time as the contemporary development of mortality approaches the mortality shifting the tempo adjusted mean age at death based on the rates of the 1st kind approaches the other tempo-adjusted measures (mean age at death based on the reduced rates or measures based on the original or simplified Kohler and Philipov methodology).

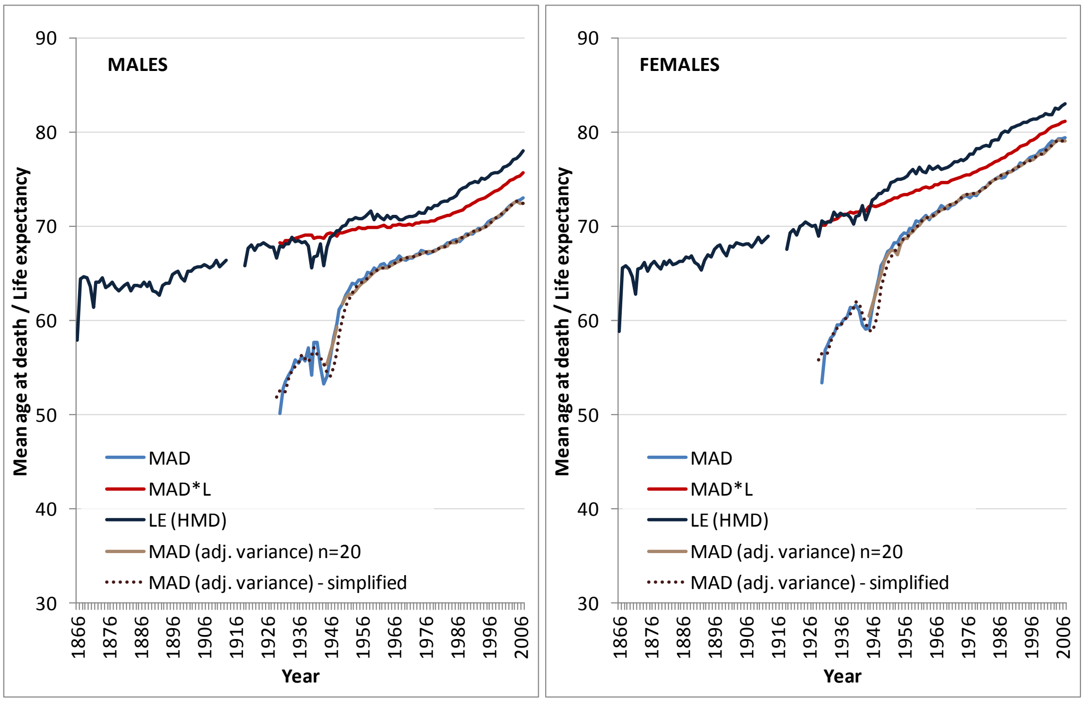
Figure 135: Life expectancy at birth compared with the tempo-adjusted mean age at death based on the rates of the 1st and 2nd kind and indicators based on the Kohler-Philipov methodology, Norway, by sex, 1866–2006



Note: "MAD" is the adjusted mean age at death based on the rates of the 2nd kind; "MAD*L" is the adjusted mean age at death based on the rates of the 1st kind; "LE(HMD)" represents the life expectancy at age 30 from the Human Mortality Database, "MAD (adj. variance) n=20" represents the mean age adjusted according to the Kohler and Philipov (2001) methodology where 20 iterations were used, "MAD (adj. variance) – simplified" is its simplified variant suggested in this chapter for the mortality process.

Source of data: author's calculation based on Human Mortality Database (2010)

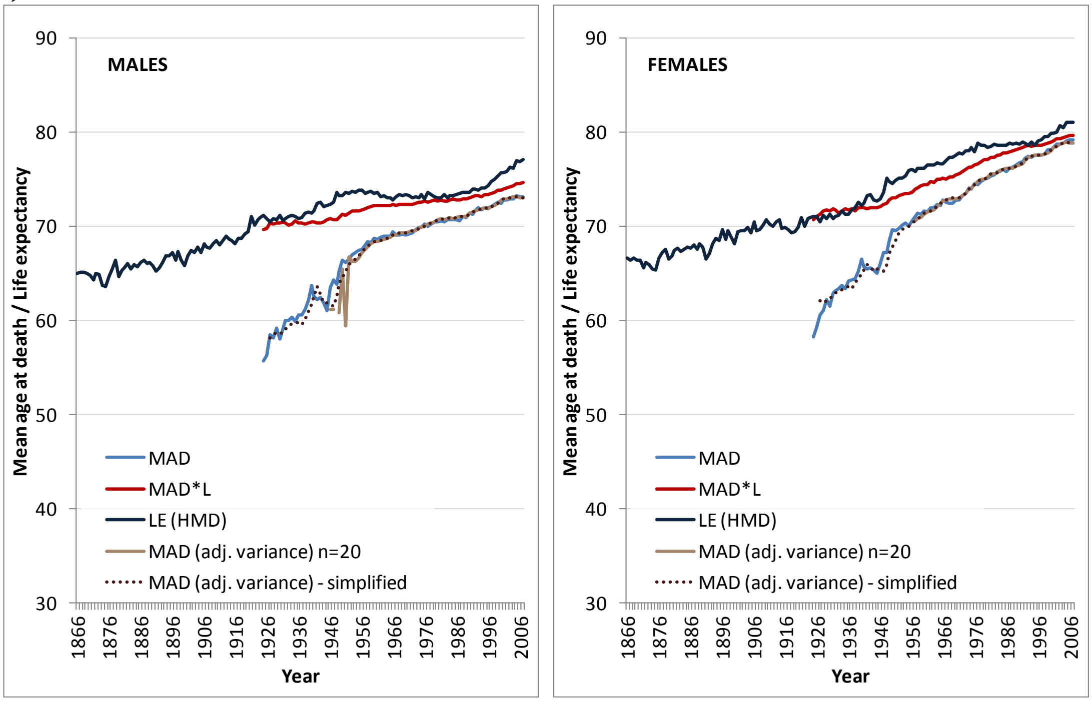
Figure 136: Life expectancy at birth compared with the tempo-adjusted mean age at death based on the rates of the 1st and 2nd kind and indicators based on the Kohler-Philipov methodology, Belgium, by sex, 1866–2006



Note: see notes for the Figure 135

Source of data: author's calculation based on Human Mortality Database (2010)

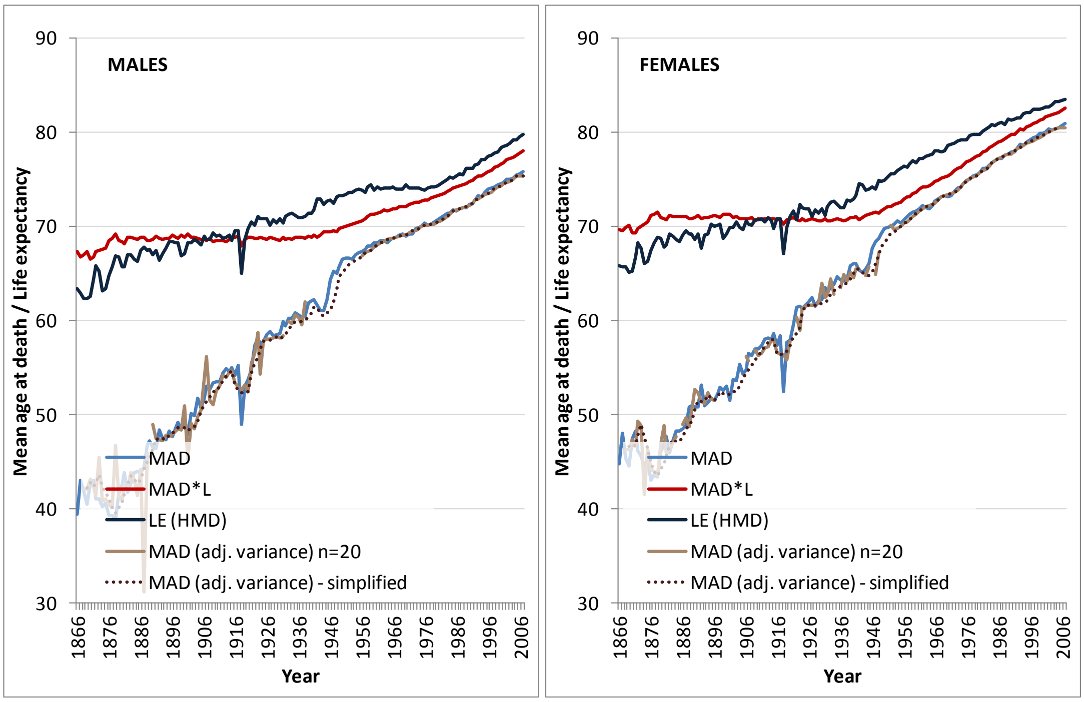
Figure 137: Life expectancy at birth compared with the tempo-adjusted mean age at death based on the rates of the 1st and 2nd kind and indicators based on the Kohler-Philipov methodology, Denmark, by sex, 1866–2006



Note: see notes for the Figure 135

Source of data: author's calculation based on Human Mortality Database (2010)

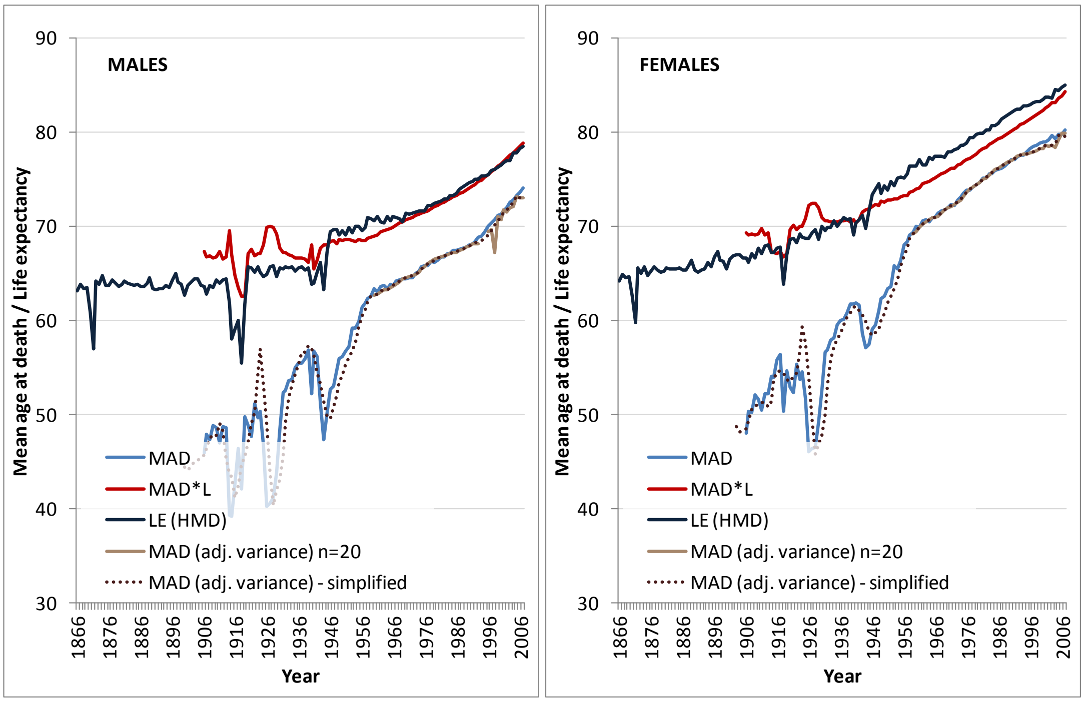
Figure 138: Life expectancy at birth compared with the tempo-adjusted mean age at death based on the rates of the 1st and 2nd kind and indicators based on the Kohler-Philipov methodology, Sweden, by sex, 1866–2006



Note: see notes for the Figure 135

Source of data: author's calculation based on Human Mortality Database (2010)

Figure 139: Life expectancy at birth compared with the tempo-adjusted mean age at death based on the rates of the 1st and 2nd kind and indicators based on the Kohler-Philipov methodology, France, by sex, 1866–2006



Note: see notes for the Figure 135

Source of data: author's calculation based on Human Mortality Database (2010)

As was shown it is possible to assume that the tempo-adjusted mean age at death (*MAD*) proposed by Bongaarts and Feeney is not distorted by the changing variability. So it could be taken as a measure of mortality development instead of the traditional life expectancy. Through this measure similar general patterns could be seen (Figure 140). First of them is the growing tendency of its values. In the past the values were affected by some crises, e.g. the World Wars what is more visible in the case of males. Also the values are more homogeneous for females in comparison to males. However it must be said that only 5 selected countries are compared. According to the results the contemporary tempo-adjusted mean age at death is in our selected countries around 75 years for males and around 80 years for females. That would mean that the ca 4–5 years (for which the traditional life expectancy is higher than the *MAD* measure) represent the tempo distortion caused by the postponement of deaths.

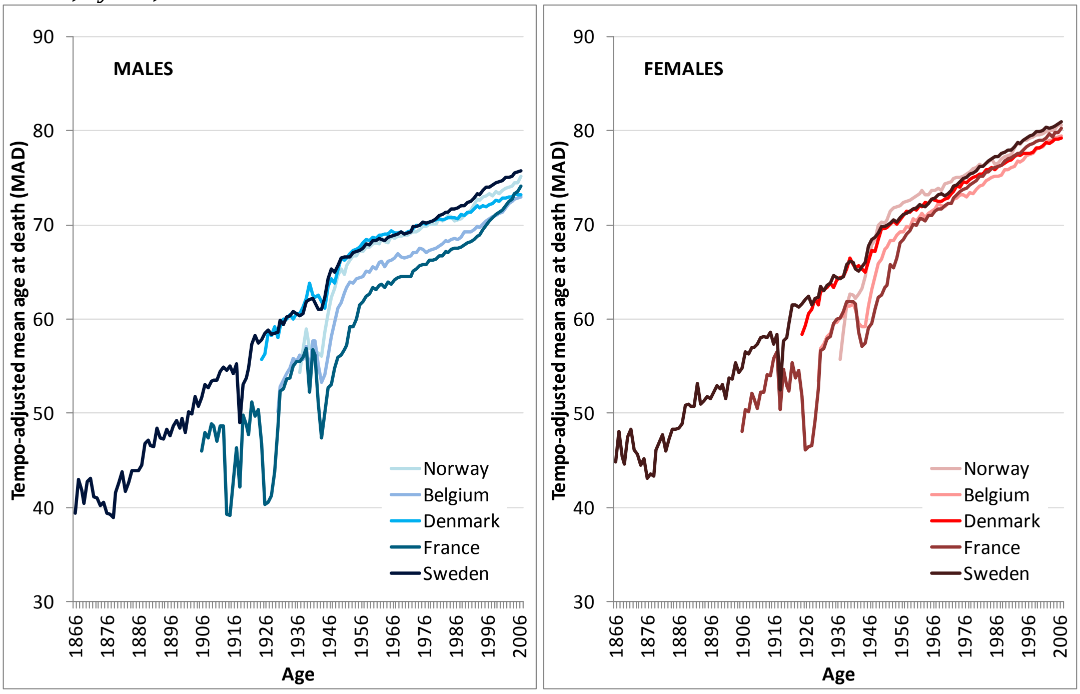
9.6 Summary

In this chapter the analysis of the tempo effect was introduced. Its application is limited by the insufficient length of the time series of the data. That was the reason why we applied the Bongaarts-Feeney's methodology to data from only 5 European countries where data are available for a long time period. The application to contemporary data (shown on the example of Sweden) did not fully confirm the results of Bogaarts and Feeney (2006). In this Thesis the estimation of the total mortality rate was slightly closer to one and also the mean age at death based on the rates of the 1st kind was more different from the mean age based on the rates of the 2nd kind. The reason lies probably in slightly different calculation of the mortality rates of the 2nd kind (reduced rates). Bongaarts and Feeney (2006) used hypothetical period and cohort life tables where mortality at ages below 30 was taken as being equal to zero. In our calculation there the numbers of deaths bellow the age 30 were subtracted from the numbers of births in corresponding cohorts because these deaths are not included also in the numerator (while the mortality below age 30 is neglected).

When the same tempo-adjusting procedure was applied to more historical data (that means to data characterized by significant changes in mortality patterns) it turned out the necessity of keeping in mind the assumption of constant shape implemented in the Bongaarts-Feeney methodology. The shape of the mortality curve changed significantly during the past (what was shown in the Chapter 8).

Through the application of the Kohler-Philipov methodology, which can handle the changing variability of the mortality schedule, it was proved that the mean age at death based on the rates of the 1st kind was probably distorted by the variance effect in the past. On the other hand the mean age at death based on the rates of the 2nd kind remains robust also under changing variability conditions. As an alternative, the mean age at death estimated by the Kohler-Philipov procedure could be used or the mean age at death estimated by the Kohler-Philipov methodology simplified for the purpose of its application to mortality data. Both those measures give almost the same results as the mean age at death based on the rates of the 2nd kind.

Figure 140: Tempo-adjusted mean age at death based on the rates of the 2nd kind, selected European countries, by sex, 1866–2006



Source of data: author's calculation based on Human Mortality Database (2010)

All the results confirm the original basic idea of the analysis of the tempo effects. That means that the traditionally used period life expectancy could be significantly distorted by the tempo effect. According to Kohler and Philipov (Kohler, Philipov, 2001) the life expectancy could be distorted also by the variance effect. The indicators of the mean age at death adjusted to one or both of those effects give almost the same results (except for the mean age at death based on the rates of the 1st kind). As a consequence of that the mean age at death based on the rates of the 2nd kind could be used for long-term comparison of the mortality development in the studied countries. The general developmental patterns were confirmed also using this indicator. According to its values (Figure 140) it could be concluded that the most favorable values of the mean age at death are reached traditionally in Sweden and Norway (from the compared countries). Values of the adjusted measures are generally lower than the traditional period life expectancy as a result of the improvements in mortality levels accompanied by the rising mean age at death in the population.

Old age isn't so bad
when you consider the alternative

Maurice Chevalier, New York Times, 9 October 1960

Chapter 10

Introduction to frailty models and theory of heterogeneity in demography[19]

10.1 Introduction and the aim of the chapter

Already from all the previous chapters it is clear that the effort to learn more about human mortality and factors influencing it is not only natural to almost all people, but also it represents one of the basic building blocks of demographic analysis and demography as a whole. Many aspects have changed in demography during its 350 years long history and development of its methods of analysis. For all the contemporary methods more often the detailed, cohort or even individual data are used. Many methods are based on the usage of modern information technologies or on the development of the original methodology and also of the theoretical background.

The aim of this chapter is to introduce briefly one of the rapidly developing methodological streams in the field of mortality analysis. This theme reached many representatives in the World's demography during the latest decades. So-called frailty models are a specific field within the survival analysis but it is often understood as an independent field of study. In the Czech demographic literature the frailty models have not been introduced until recently. First information about this concept brought Koudelka and Lustigová (Koudelka, Lustigová, 2010, pp. 237–240). The first article devoted solely to the issue of frailty models in the Czech demographic literature was prepared in accordance to this Dissertation Thesis. It was published in 2012 (Hulíková Tesárková, 2012).

Frailty models could be taken as a specific part of the more general survival analysis, where they enable to implement also the influence of random effects (Wienke, 2011; Aalen *et al.*, 2010). Survival analysis (more specifically its part focused on modeling of proportional hazards used within the frailty models) represents the influence of so-called unobserved heterogeneity to the

---

[19] This chapter is based on the article prepared independently within the grant project GA UK 0136/2010 „*Způsoby zkoumání procesu úmrtnosti se zaměřením na nejvyšší věkové skupiny*" („*Methods of mortality analysis focused on the oldest age-groups*") published in 2012 (Hulíková Tesárková, 2012). The author is also grateful for the possibility to attend the course „*Frailty Models*" in January 2011 in the *Max Planck Institute for Demographic Research, Rostock, Germany* lectured by *dr. Trifon I. Missov*.

survival function or to the overall hazard function (Wienke, 2011). The analysis of proportional hazards is based on the assumption that the particular hazard functions are different for various sub-populations but that they are in a proportional relation. Particular sub-populations are defined by a given level of some selected measurable factor. The pattern of the hazard function of the studied process is then the same for all those sub-populations but the proportionality is expressed by a multiplicative constant. This constant shows how many times the intensity is higher or lower for a particular sub-population described by a given value of the external measurable factor in comparison to some other sub-population with a different value of this factor. Usually for this comparison some reference sub-population is used. In the case of the basic type of the frailty models the principle is the same. Only for the frailty models the sub-populations are not defined by any measurable value of some external selected factor but by some non-measurable value of frailty which then characterizes the sub-populations (Vaupel *et al.*, 1979; Wienke, 2011). In such models the frailty could be understood as a discrete (it is possible to define a finite number of sub-populations) or continuous variable. When frailty is a continuous variable, then each individual in a studied population could be characterized by some concrete value from a continuous interval.

In this chapter, the concept of heterogeneity of human populations (more specifically the unobserved heterogeneity) will be briefly introduced together with the role of frailty models in the survival analysis and their discrete as well as continuous forms with a practical example and illustration of their possible usage for the real data. After the chapters devoted to many aspects of period indicators, in this, the last one, the introduced methodology represents the shift towards the cohort analysis. Frailty models are primarily intended for cohort data but as it will be shown its application to period data could be also useful. That is the reason why this issue was included to the Thesis and why it was put to the end.

10.2 Theoretical background

The survival analysis is a method within the mortality analysis which is recently used still more often, but not only in the field of mortality analysis. Application of the survival analysis is possible in all the fields of demography and also in other fields of study. The aim is to study the time of transition from the initial event to the occurrence of some other defined event (for example the time from birth to death or the time from the beginning of the study to death or also the time between the marriage and divorce or the birth of the first child, the time from the medical surgery to the appearance of complications or to the state of complete health of the patient, the time from the implementation of some machine to the industry to its first failure, etc. For other examples see Hougaard, 1999, s. 13). In general in the survival analysis we consider the time until the occurrence of some defined event which will or will not meet all the studied individuals. All the measured time durations are then described by the survival curves or by the hazard functions together with the possible effect of considered explaining variables (Aalen *et al.*, 2010).

First of all it should be clarified what is the specific feature of the survival analysis, which is logically reflected also in its methods. The apparent particularity of the survival analysis is in the character of required data. One of the specifics of this data is the censoring – there is not all the information available for the censored data (Hougaard, 1999, p. 15). The censoring means that for example we know that a particular person died or will die in the future, but we do not know when this event occurred or will occur. That person could have got lost from the study or the studied event did not have to occur until the end of the study – the person did not die before the end of the study or he or she died because of some other cause than the studied one (Hougaard, 1999; Aalen *et al.*, 2010). Another important characteristics of the data used in the survival analysis (for some authors, i.e. Hougaard, 1999, p. 15, even more important than censoring) is the so-called truncation (Aalen *et al.*, 2010) or conditioning (Hougaard, 1999) or left censoring (Aalen *et al.*, 2010). The principle of truncation is the fact that not all the studied persons enter the study at the same age (or after the same time from the initial event, in general). This fact could be illustrated by some simple example: the probability of surviving or dying during some time or age interval is different when it is studied as unconditional or conditional. The probability of survival to the age of 85 is different for a just born person in comparison with a person under some condition, for example surviving to age 84. Similar example is presented also by Hougaard (1999, p. 15). The left censored data are sometimes called also as data with delayed entry (Aalen *et al.*, 2010, p. 5). The practical impact for the studied data set is the fact that the number of persons exposured to the risk does not have to decline during the study (as a consequence of the occurrence of the studied event) but it can also increase because of the delayed entries (Aalen *et al.*, 2010). Another important fact is that the traditionally used parametric methods do not always respect the character and development of the analyzed data. That is the reason why within the survival analysis nonparametric methods are so popular (Hougaard, 1999).

One of the most common formulations of the model in the survival analysis includes also the influence of some known explaining variables. Then according to for example Hougaard (1999), Aalen *et al.* (2010) or Wienke (2011, p. 43) it is possible to write:

$$\mu(x|A) = \mu_0(x) * h\,(A),$$

where $\mu(x|A)$ is the hazard function of the studied event (the conditional probability of its occurrence in time or at the age x[20] under the condition that the event have not occurred before that moment) for the age x and an array of explaining variables A, $\mu_0(x)$ is the value of the baseline hazard function and $h\,(A)$ is a positive function of the explaining variables. Very often this function is defined as

$$h(A) = e^{\beta` A},$$

where $\beta`$ is the transposed array of parameters representing the influence of the explaining variables[21]. The model in the presented form (referred to also as the Cox model or Cox

[20] In this chapter the symbol „*x*" stands for age as well as for the time duration from the initial event because also the age could be taken as the time from the moment of birth. The symbol „*x*" was selected because the standard practice in the field of demographic analysis of mortality.

[21] An alternative way of expression of the model is the Aalen's formula in the form (Aalen *et al*., 2010, p. 8).

regression) can be estimated in such a way that it is not needed to cope with the unknown pattern of the baseline hazard, it is possible to focus only on the study of the influence of the explanatory variables (Cox, 1972). The presented form of the model is based on the assumption of proportional intensities for all the studied persons (Aalen *et al.*, 2010).

The simplest models of the survival analysis work with the assumption of independent and identical data distribution as a consequence of homogeneity of the population with respect to differences according to included variables (Wienke, 2011). The above mentioned formulas could be used in the survival analysis for the study of the influence of selected measurable variables to the survival or hazard function.

10.3 Unobserved heterogeneity and frailty models – discussion

It is possible to study not only the measurable and observable variables but also the so-called unobserved (hidden) heterogeneity. The observable and measurable heterogeneity represents the influence of factors like the age, sex, marital status, or education attainment, etc. On the other hand, the unobserved heterogeneity can include for example the influence of the environment, genetic factors, individual life style or inborn predispositions. These non-measureable factors are the most frequent source of variability in the usually used models. Although demographers were always aware of this unobserved and non-measurable heterogeneity, finding of some methods which can deal with it was much more complicated.

The first who was concerned with this issue was Beard (1959). He studied the aspects of the assumption that the whole population could be stratified according to a specific variable which he called the "longevity factor". He derived that the intensity of mortality for the whole population is in fact the weighted average of particular intensities of the sub-populations with a concrete value of the longevity factor.

The first study introducing the term "frailty models" is from the year 1979 (Vaupel *et al.*, 1979). This study started the rapid development of demographic analysis in this field. The influence of the unobserved heterogeneity and its possible consequences for the mortality analysis are clearly introduced also by Vaupel and Yashin (1985). Their article could be used also as a methodological base when dealing with the so-called discrete frailty models (see below).

The question, whether the frailty or a predisposition to death are constant during the life (are inborn) or could change (as a result of the influence of some external factors) was solved for example by Yashin *et al.* (1994). They concluded that (under some conditions[22]) models with constant frailty during the life are not only distinguishable from models with stochastically changing frailty but also that they give the same results. Various types of probabilistic distribution usable for frailty modeling in the continuous case (see below) were studied also by Steinsaltz and Wachter (2006).

$\alpha(x|a_1,\ldots,a_p) = \beta_0(x) + \beta_1(x) * a_1 + \cdots + \beta_p(x) * a_p$, where $\beta_0(x)$ is again the baseline hazard function and the regression functions $\beta_j(x)$ describe the relationship with the explaining variables $a_1 \ldots a_p$ in time or at the age x (*ibid.*).

[22] For example the assumption of validity of the Gamma-Makeham mortality model with constant frailty or Le Bras model with stochastically changing frailty.

The fundamental monograph dealing with the concept of frailty models and their connection with the survival analysis as well as their further development and enrichment (shared or correlated frailty models) is the publication of Andreas Wienke (2011). The theory and application of this concept was not only in the past but it is also today connected mainly with the Max Planck Institute for Demographic Research in Rostock, Germany[23].

10.4 Illustration of the unobserved heterogeneity

Some simple example of including of the unobserved heterogeneity into the model could be shown for a hypothetical situation. We can suppose that there are only two relatively homogeneous sub-populations in the overall population. The mortality level in one of them is higher than in the other. The force (intensity) of mortality (hazard function) for the first sub-population could be marked in accordance to the cited literature (Wienke, 2011) as $\mu_1(x)$ and for the second sub-population as $\mu_2(x)$ where *x* is age as usually. Then the total observed hazard function (intensity of mortality) for the whole population is $\mu(x)$. We can also assume for example that the proportion of both those sub-populations at the moment of birth is equal. As time goes by, the individuals in both the sub-populations grow older. Logically those whose intensity of mortality is higher die at on average lower age. Thereof the proportion of the sub-population with higher mortality is decreasing with age and on the other hand the proportion of the sub-population with lower mortality goes up. The proportion of survivors to the age *x* who are in the first sub-population could be marked as $\pi_1(x)$ and the proportion of those who are in the second sub-population as $\pi_2(x)$. Then holds

$$\pi_2(x) = 1 - \pi_1(x).$$

It could be written (Vaupel, Yashin, 1985, p. 176):

$$\pi_1(x) = \frac{\pi_1(0)*l_1(x)}{\{\pi_1(0)*l_1(x)+[1-\pi_1(0)]*l_2(x)\}}$$

where $l_i(x)$ stands for the proportion of survivors to the age *x* in the sub-population *i*, that means it is the survival function from the life table where the $l_i(0)$, the radix of the table, could be for simplicity taken as being equal to 1. The survival function for a given sub-population (and because the radix is equal to one it is also the proportion of survivors to the age *x* in that sub-population) could be written as (*ibid.*):

$$l_i(x) = e^{-\int_0^x \mu_i(t)dt},$$

where *i = 1, 2* for the two particular sub-populations.

The intensity of mortality observable for the whole population, $\mu(x)$, is then the weighted average of intensities in both sub-populations as was defined already by Beard (1959) or Vaupel *et al.* (1979) and Vaupel, Yashin (1985, p. 176):

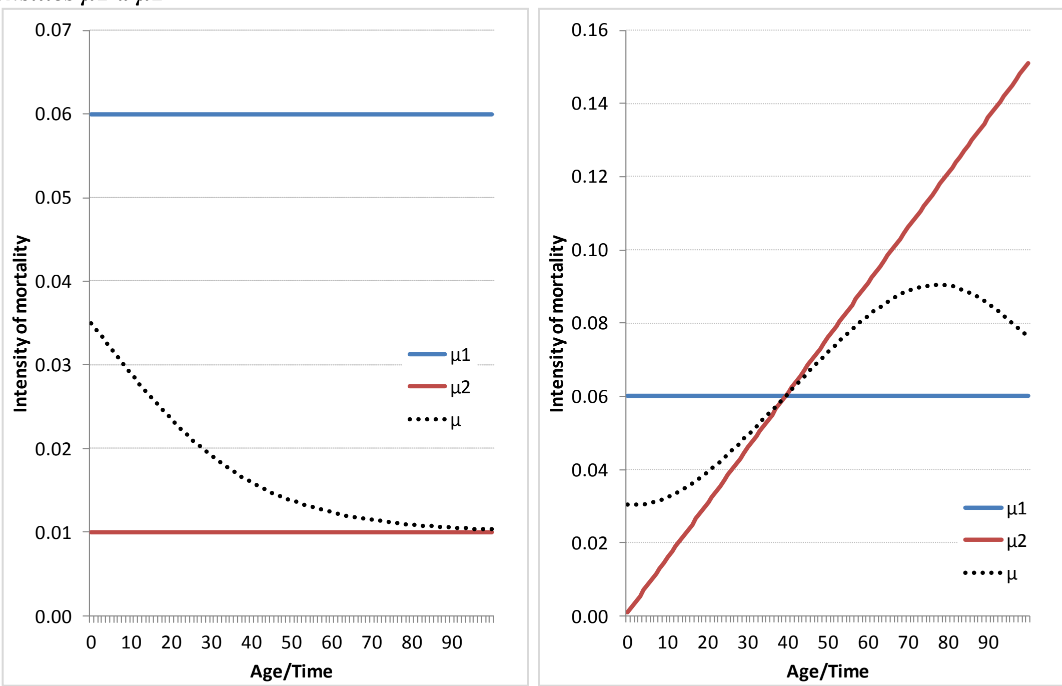
[23] http://www.demogr.mpg.de/

$$\mu(x) = \pi_1(x) * \mu_1(x) + [1 - \pi_1(x)] * \mu_2(x).$$

As was shown for example by Vaupel and Yashin (1985) and in the Czech demographic literature by Koudelka and Lustigová (2010) or Hulíková Tesárková (2012), the intensity of mortality observable for the whole population could be significantly different from the pattern of the intensity for each particular sub-population. It is the consequence of changing proportions of those sub-populations in the population as a whole caused by dying of individuals with higher level of mortality (i.e. with higher level of frailty). With increasing age, mortality of the whole population approaches mortality of the predominant sub-population (Vaupel, Yashin, 1985). According to some theories, a possible consequence of this fact could be also the existence of the so-called mortality crossovers when patterns of several sub-populations intersect each other (Carey *et al.*, 1995). The examples shown in the Figures 141 and 142 are taken from the paper of Vaupel and Yashin (1985) and are only modified for the needs of this Thesis.

From the examples it is clear that also relatively complicated patterns of the studied functions could be expressed as a mixture of simpler functions. This could be also the way to a solution of a typical demographic problem within mortality analysis when many demographers tried to find suitable analytical functions reflecting as good as possible the mortality pattern with age (see for example Burcin *et al.*, 2010 or Chapter 5).

Figure 141: Model example of the overall intensity of mortality, observable for the whole population (μ) under the assumption that the total population is formed by two homogeneous sub-populations with intensities μ1 a μ2

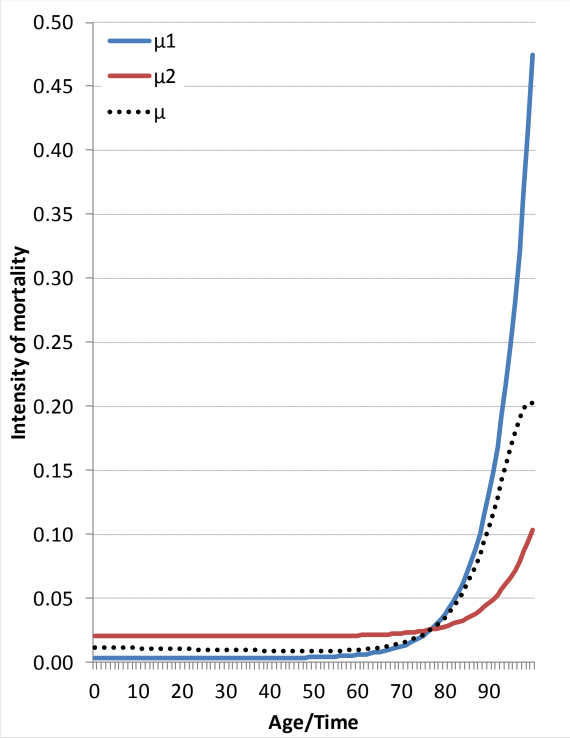


Note: μ1 = 0.06 (both examples) a μ2 = 0.01 (on the left side) and μ2 = 0.0010 + 0.0015 * x (on the right side), where x stands for age / time. The proportion of both sub-populations at age / in time x = 0 is equal to 50 %.

For illustration we can imagine a mortality function represented by the classical Gompertz-Makeham relation for example in the form

$$\mu(x) = a + b * c^x,$$

Figure 142: Model example of the overall intensity of mortality, observable for the whole population (μ) under the assumption that the total population is formed by two homogeneous sub-populations with intensities μ1 a μ2 corresponding to Gompertz-Makeham function



Note: μ1, μ2 = a + b * c^x, where x stands for age / time. Parameter values for μ1 are taken as: a = 0.003, b = 9 * 10^{-7}, c = 1.14, for μ2 as: a = 0.02, b = 6 * 10^{-7}, c = 1.125. The proportion of both the sub-populations at age / in time x = 0 is equal to 50 %.

where *a*, *b*, *c* are the parameters of the function and *x* stands again for age. This law represents the exponentially increasing mortality with age and recently it is sometimes also criticized for overestimating the empirical data at the highest ages and often it is replaced by the logistic curve which enables to include also the decreasing tempo of the mortality increase with age at those highest age groups (Burcin *et al.*, 2010). However, it is possible to show that in the case when the Gompertz-Makeham function is considered for particular sub-populations the overall mortality (observable for the whole population) could follow by its shape the empirical data very well and even it is approaching the pattern of a logistic function (Figure 142). The analytical prove of this fact presents for example Beard (1959).

10.5 Discrete frailty models

In the previous part of this chapter in fact the discrete frailty models were already shown and formulated. They are based on the fact that we consider a given (finite) number of relatively homogeneous sub-populations which differ according to the resistance of the individuals. This resistance will be considered as constant during the life (so it represents some inborn predispositions) (Yashin *et at.*, 1994).

In the discrete frailty models we consider a certain number of sub-populations which differ according to its level of frailty. The level of frailty is then represented by a multiplicative constant for each of those sub-populations. The overall intensity of mortality is then (as was already mentioned) the weighted average of all the particular intensities. Although this type of models (in the discrete form) could look like relatively simple for its estimation, its important disadvantage is the total number of unknown parameters which is relatively high. Except for the parameters of the baseline hazard function it is necessary to estimate the number of sub-populations, their proportion at the beginning of the study (or in the moment of birth), then also for each sub-population the level of frailty should be estimated (except for one sub-population which could be taken as a reference group where the level of frailty could be equal to one and so the intensity of this sub-population could be equal to the baseline intensity). A model defined in this way is already relatively complicated for estimation, so for simplicity it is possible to choose the number of sub-populations ex ante and to estimate only the parameters of the baseline hazard and particular levels of frailty. The calculation could be repeated for different numbers of sub-populations and all the models and results could be then compared and evaluated with respect to the statistical significance of each added sub-population.

10.6 Continuous frailty models

The continuous frailty could be imagined as a generalization of the discrete case when the number of considered sub-populations is increasing until each of them contains only one individual person form the studied population. In that moment the frailty becomes a random variable with certain continuously defined distribution in the total population and represents characteristics for each individual.

The continuously defined model is paradoxically easier for estimation in comparison to the discrete one. It is not necessary to estimate the number of sub-populations neither the individual values of frailty in those sub-populations. Frailty, as a random variable, could be expressed by a suitable statistical distribution. The estimation of the model is then limited only to the estimation of the parameters of the baseline hazard and parameters of the frailty distribution. So that the estimation is easier and the results simply interpretable, usually some other assumptions are also incorporated: the frailty could reach only positive values, frailty distribution is the same for all the individuals in the population and the average value of frailty at the beginning of the study (for $x = 0$) is equal to one (Wienke, 2011).

A model defined as above could be expressed on the basis of the concept of proportional hazards as (Wienke, 2011, p. 57; Vaupel *et al.*, 1979, pp. 440–441):

$$\mu(x|Z) = Z * \mu_0(x),$$

where $\mu(x|Z)$ represents the intensity of mortality for an individual in time (or at age) x with frailty Z and $\mu_0(x)$ is the baseline hazard[24]. From the equation it is clear that the intensity of mortality for an individual with frailty $Z = 1$ is equal to the baseline hazard $\mu_0(x)$. If the assumption that the expected value of frailty (EZ) in the moment of the beginning of the analysis ($x = 0$) is equal to one ($EZ = 1$) is accepted, then the baseline mortality represents the intensity of the average individual. The variance of the frailty distribution could be taken as an indicator of the heterogeneity in the studied population (Wienke, 2011). For such a simply defined model the survival function could be expressed in the form (Wienke, 2011, p. 59):

$$S(x|Z) = e^{-\int_0^x \mu(t|Z)dt} = e^{-Z*M_0(x)},$$

where $M_0(x)$ is the cumulative baseline hazard (the cumulated conditional probability of the occurrence of the event),

$$M_0(x) = \int_0^x \mu_0(t)dt.$$

[24] In case of existence of some other variables representing measurable factors this model transforms to the form (Wienke, 2011): $\mu(x|A,Z) = Z * \mu_0(x)e^{\beta`A}$.

As the defined hazard function so the survival function, both are so-called conditional functions, that means the individual expressions of those functions (expressed for an individual with some concrete value of frailty). In fact, those are not for a standard real population measurable. Therefore, within the model it is necessary to focus only on formulas which do not include the unknown and non-measurable frailty, so to focus on formulas defined for the whole population where data are available. We have to return to the assumption stated above that the survival function of the population as a whole is the weighted average of the individual functions (*ibid*.)

$$S(x) = E\, S(x|Z) = E\, e^{-Z*M_0(x)} = L(M_0(x)),$$

where L is the Laplace transform known from mathematics or statistics[25]. The usage of the Laplace transform in these models enables to simplify the model and its further estimations[26].

However, still there remains a question of the appropriate choice of the frailty distribution in population. It is important to decide about the distribution at the beginning of the study (or at the time of birth, or in time $x = 0$, in general). Although we consider the frailty as invariant during the life, the composition of the population according to the level of frailty changes because the more frail individuals die on average at lower ages. As a consequence of that, the distribution of frailty in the population changes. Again it holds that the overall intensity of mortality observable in the population is the weighted average of individual intensities where the weights are defined by the frailty distribution in the population. In general, as the frailer individuals die the average level of frailty in the overall studied population decreases. From that follows that the individual intensity increases with age more rapidly than the overall intensity observable in the population as a whole (Wienke, 2011; Vaupel *et al.*, 1979). From these facts Vaupel *et al.* (1979) derived why on average individuals grow older more rapidly than the whole population or cohort.

From the technical point of view the most frequently and successfully used distribution of frailty in the population is the Gamma distribution (Wienke, 2011; Vaupel *et al.*, 1979; Beard, 1959). There are many advantages of this choice:

1) Random variable from this distribution reaches only positive values;
2) If the expected value of the random variable Z is taken as to be equal to one then the Gamma distribution is reduced to a distribution with only one unknown parameter;
3) The Laplace transform is easy usable for the Gamma distribution;
4) The Gamma distribution is flexible enough – it could be significantly asymmetrical or also almost symmetrical similar to the normal distribution. It is important because thanks to that we do not need to decide about any restricting assumptions for the distribution before the estimation of the model.

Not only the Gamma distribution, but also for example the inverse Gauss (normal) distribution or some other statistical distributions are used frequently (Wienke, 2011; Steinsaltz, Wachter, 2006).

---

[25] The Laplace transform is one of the basic mathematical transforms and it is defined (e.g. according to Wienke, 2011) as: $L(u) = Ee^{-uT} = \int_0^{\infty} e^{-ut} * f(t)dt$, where T is the random variable and $f(t)$ is its probability density.

[26] Other derived equations for the functions of random variable are presented e.g. in (Wienke, 2011).

Except for the choice of the statistical distribution for the frailty it is necessary to deal with the choice of the distribution of the baseline hazard, as it is the traditional problem in the demographic mortality analysis solved in the process of life table construction (Burcin *et al.*, 2010 or Chapter 5 of this Thesis). In this case, as was already shown above, it is not necessary to look for any relatively complicated models because the mixture of more simple models while incorporating the frailty enables to reach the resulting pattern well corresponding with the empirical data. Another possibility is the usage of the survival analysis methodology and non-parametric expressions of mortality (for more information about these methods see e.g. Aalen *et al.*, 2010 or with the connection to the frailty models see Wienke, 2011).

10.7 Application of frailty models to real data for the Czech Republic

This type of models as well as the whole survival analysis is developed mainly for the cohort data because it contains also modeling of the changes in the cohort of individuals with concrete frailty of each of them and studies their progressive extortion (or other defined exit from the studied population). It could be taken as a limitation of these methods from the data-availability point of view. It is possible to use data from survey samples or some short-term studies. If we would like to study mortality using the cohort data we will find out that in the Czech Republic there are no published data about mortality of whole cohorts from their birth to the death of the last one individual. For the basic illustration of the usage of frailty models in the mortality analysis some simplification will be done in this Doctoral Thesis[27]. We will use the period data (the population of persons deceased during the year 2009 in the Czech Republic). All the people in the studied data set will represent individuals of one hypothetical cohort (although in real they came from ca 100 various cohorts). Thanks to the data about their individual lengths of life it is possible to follow the extortion of this model fictitious cohort[28]. It is a significant simplification of the assumptions of frailty models and the overall survival analysis. But on the other side it should be remembered that almost the same simplification is used for all the life tables' constructions. To the model analysis then 54 080 males and 53 341 females entered.

For illustration it is possible to use probably the most frequent combination of distributions. The Gompertz function with two unknown parameters α and β will be used for the baseline hazard function in the form

$$\mu_0(x) = \alpha * e^{\beta * x}.$$

The cumulative baseline hazard of this distribution has the form:

$$M_0(x) = \int_0^x \alpha * e^{\beta * t} dt = \alpha * \int_0^x e^{\beta * t} dt = \frac{\alpha}{\beta}\left[e^{\beta * t}\right]_0^x = \frac{\alpha}{\beta} * \left(e^{\beta * x} - 1\right)$$

[27] The same example was used also in the already published article (Hulíková Tesárková, 2012).
[28] The same data set was used already in the Chapter 6.

For the frailty the Gamma distribution will be used (we define it by its probability density) with parameters λ and k, where $\Gamma(k)$ is the gamma function of k:

$$f(z) = \frac{1}{\Gamma(k)} * \lambda^k * z^{k-1} * e^{-\lambda * z}.$$

If we hold the assumption that the expected value of frailty at the beginning of the study (when $x = 0$) is equal to one then $\lambda = k$ and the Gamma distribution could be defined with only one parameter (k):

$$f(z) = \frac{1}{\Gamma(k)} * k^k * z^{k-1} * e^{-k * z}.$$

For such a defined model the intensity of mortality of the whole population would have the form:

$$\mu(x) = \frac{\alpha * e^{\beta * x}}{1 + \frac{1}{k} * \frac{\alpha}{\beta} * (e^{\beta * x} - 1)},$$

where α, β and k are the parameters, the first two parameters are related with the baseline hazard and model the relation of mortality and age (the Gompertz function), the third parameter is related to the frailty distribution in the population (the Gamma distribution) where this parameter characterizes the level of heterogeneity of the studied population. Again x refers to the time from the beginning of the study, in this concrete example it is the age. This above mentioned form of the model is introduced also by Wienke (2011, p. 76) and can be simply derived as follows:

The baseline distribution is the Gompertz function with parameters α and β. The form of this distribution is fairly known:

$$\mu_0(x) = \alpha * e^{\beta * x}.$$

Then the cumulative baseline hazard is:

$$M_0(x) = \int_0^x \alpha * e^{\beta * t} dt = \alpha * \int_0^x e^{\beta * t} dt = \frac{\alpha}{\beta} \left[e^{\beta * t} \right]_0^x = \frac{\alpha}{\beta} * \left(e^{\beta * x} - 1 \right).$$

For the frailty the Gamma distribution was used in the Gamma $\Gamma(k,k)$[29] form:

$$f(z) = \frac{1}{\Gamma(k)} * k^k * z^{k-1} * e^{-k * z}.$$

The Laplace transform for the Gamma distribution is:

[29] $\Gamma(k)$ stands for the so-called Gamma function of the value k; $\Gamma(k, k)$ stands for the Gamma distribution with both parameters equal to k.

$$L(u) = \frac{1}{\Gamma(k)} * k^k * \int_0^\infty e^{-u*z} * z^{k-1} * e^{-k*z} dz$$

$$= \frac{k^k}{(k+u)^k} * \frac{1}{\Gamma(k)} * (k+u)^k * \int_0^\infty z^{k-1} * e^{-(k+u)*z} dz = \left(1 + \frac{u}{k}\right)^{-k}$$

Then the derivation of the Laplace transform could be written as:

$$L'(u) = \frac{-k}{k} * \left(1 + \frac{u}{k}\right)^{-k-1} = -\left(1 + \frac{u}{k}\right)^{-k-1}$$

The unconditional survival function for the whole population could be derived in a form:

$$S(x) = E\ S(x|z) = E\ e^{-z*M_0(x)} = L\big(M_0(x)\big) = \left(1 + \frac{M_0(x)}{k}\right)^{-k},$$

where $M_0(x) = \frac{\alpha}{\beta} * \left(e^{\beta*x} - 1\right)$, so:

$$S(x) = \left[1 + \frac{1}{k} * \frac{\alpha}{\beta} * \left(e^{\beta*x} - 1\right)\right]^{-k}.$$

The unconditional probability density could be written as:

$$f(x) = -\mu_0(x) * L'\big(M_0(x)\big) = -\alpha * e^{\beta*x} * \left[-\left(1 + \frac{M_0(x)}{k}\right)^{-k-1}\right] =$$

$$\alpha * e^{\beta*x} * (1 + \frac{M_0(x)}{k})^{-k-1},$$

where $M_0(x) = \frac{\alpha}{\beta} * \left(e^{\beta*x} - 1\right)$, so:

$$f(x) = \alpha * e^{\beta*x} * (1 + \frac{1}{k} * \frac{\alpha}{\beta} * \left(e^{\beta*x} - 1\right))^{-k-1} = \alpha * e^{\beta*x} * \frac{1}{\left[1 + \frac{1}{k} * \frac{\alpha}{\beta} * \left(e^{\beta*x} - 1\right)\right]^{k+1}}.$$

Finally the intensity of mortality of the total population (which is the ratio of the probability density and survival function) has the form:

$$\mu(x) = -\mu_0(x) * \frac{L'\big(M_0(x)\big)}{L\big(M_0(x)\big)} = -\mu_0(x) * \frac{-\left(1 + \frac{M_0(x)}{k}\right)^{-k-1}}{\left(1 + \frac{M_0(x)}{k}\right)^{-k}}$$

$$= \mu_0(x) * \left(1 + \frac{M_0(x)}{k}\right)^{-k-1-(-k)} = \mu_0(x) * \left(1 + \frac{M_0(x)}{k}\right)^{-1} = \mu_0(x) * \frac{1}{1 + \frac{1}{k} * M_0(x)}$$

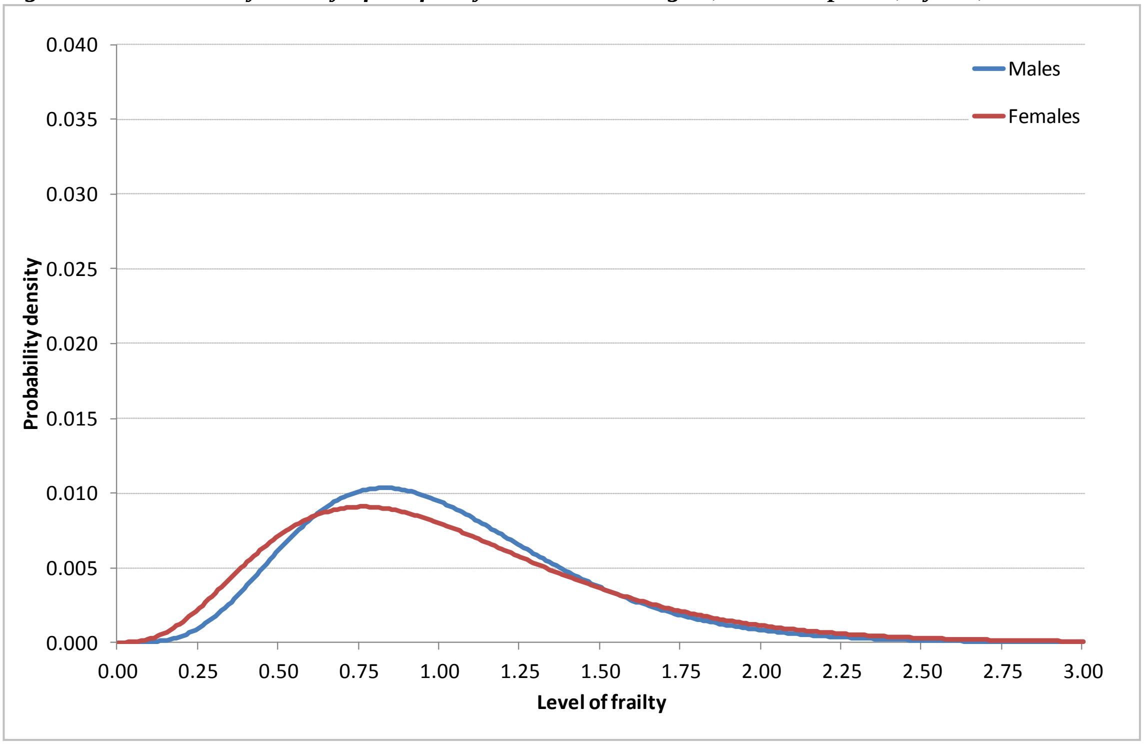
$$= \mu_0(x) * \frac{1}{1+\frac{1}{k}*\frac{\alpha}{\beta}*(e^{\beta*x}-1)} = \frac{\alpha*e^{\beta*x}}{1+\frac{1}{k}*\frac{\alpha}{\beta}*(e^{\beta*x}-1)}$$

The estimation of the three unknown parameters was made in the R software (R Development Core Team, 2009) and the method of maximal likelihood was used.

According to the results, parameters of the Gompertz function were relatively similar for males and females, the parameter α was for males equal to $2.69*10^{-8}$ and for females $3.98*10^{-9}$, parameter β was estimated as being equal to $2.96*10^{-1}$ resp. $2.47*10^{-1}$. The estimation of the third parameter k is also the estimation of the reciprocal value of the level of the unobserved heterogeneity in the studied population (Vaupel *et al.*, 1979, p. 443). In our case the unobserved heterogeneity was lower in case of males (the value of the parameter was 5.74 in comparison to females where the value of the parameter was 4.14).

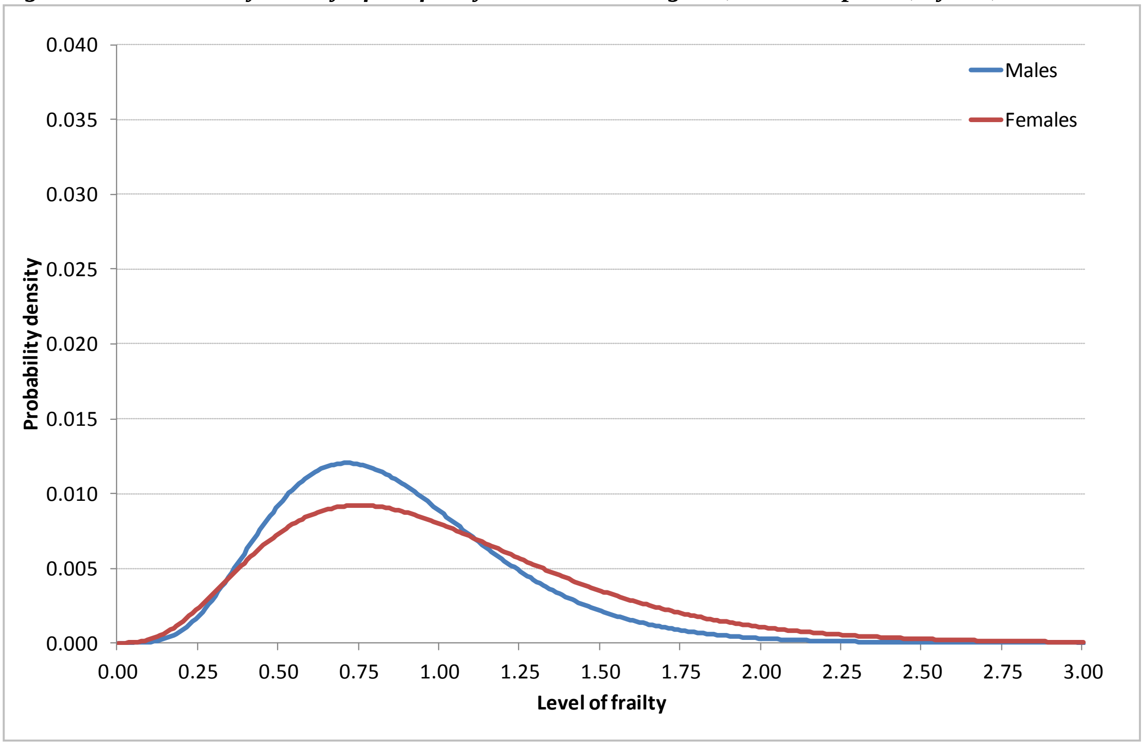
All the results are presented here in the form of graphs. From the results it could be seen that the average level of frailty of the survivors decreases with age for males as well as for females. But for males the decrease occurs at lower age, so at relatively lower age the less resistant males die in comparison to females with the same level of frailty. While the surviving males reach the average frailty equal to 0.5 around the age 70, females reach this average level of frailty some 10 years later. In other words, an individual with the same level of frailty lives some 10 years longer if it is a female in comparison to the situation that it is a male.

Figure 143: Probability density of the frailty distribution at age 0, Czech Republic, by sex, 2009



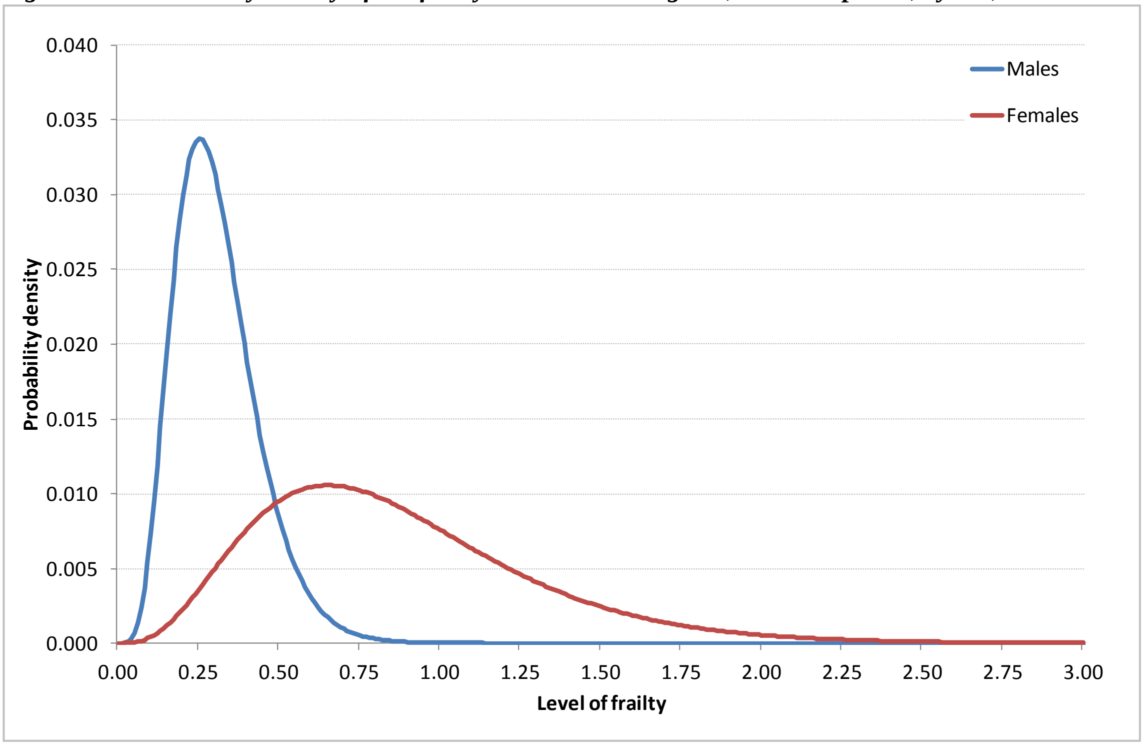
Source of data: author’s calculation based on Czech Statistical Office (2011a)

Figure 144: Probability density of the frailty distribution at age 60, Czech Republic, by sex, 2009



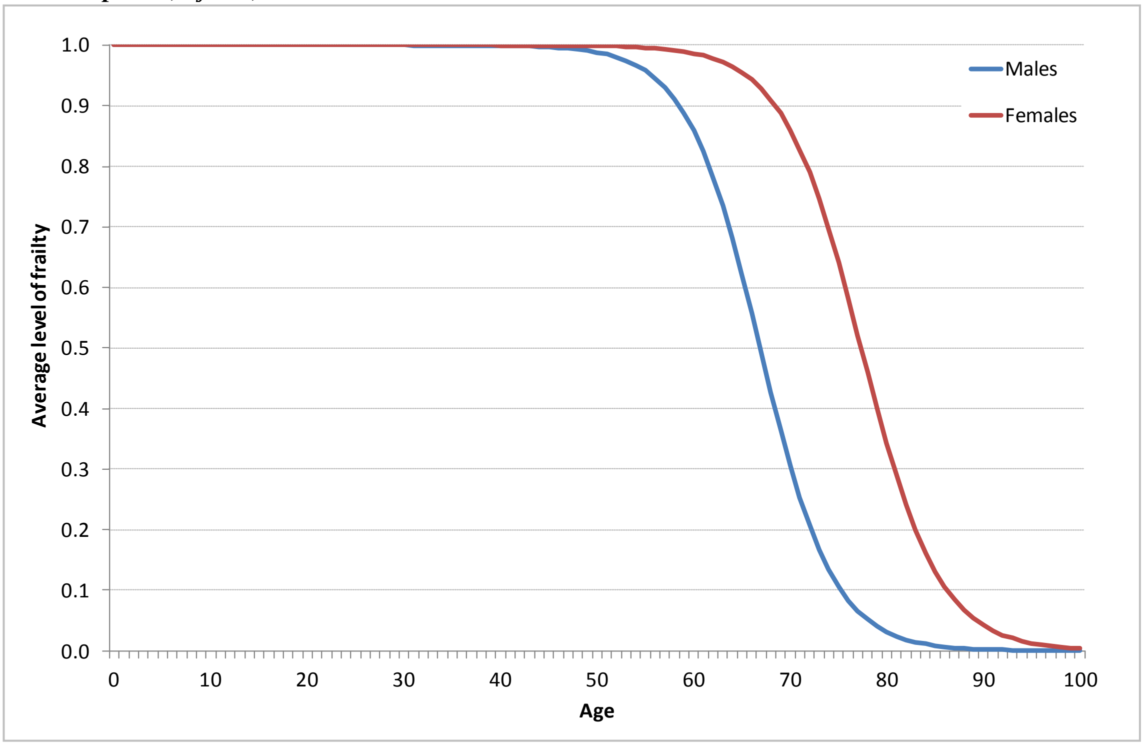
Source of data: author's calculation based on Czech Statistical Office (2011a)

Figure 145: Probability density of the frailty distribution at age 70, Czech Republic, by sex, 2009



Source of data: author's calculation based on Czech Statistical Office (2011a)

Figure 146: Relation of the average level of frailty in the population and the age of survivors, Czech Republic, by sex, 2009



Source of data: author's calculation based on Czech Statistical Office (2011a)

It is also clear that for males as well as for females only a small group of the most resistant (the least frail) individuals from the total population survive to the age of 100. But it is necessary to keep in mind that the calculation of this model example was done on a fictitious cohort formed from all the persons deceased during one calendar year (not on a real cohort). Those people were born during many years and they were members of many (nor equally numerous) cohorts. These facts surely influence the results.

10.8 Conclusions

In this chapter only the basics of the frailty models were introduced because the detailed description of this concept would take the range of the whole Thesis. Briefly also the theoretical background of the frailty models as a special method of the survival analysis was mentioned. These methods are still more and more popular in demography because also the data are more available and in a better quality and also the information technologies develop. It was shown here that only the simple mixture of several sub-populations with the same mortality pattern (the same function but different values of the parameters) could model relatively complicated patterns observable for the whole population. The estimation of such formulated discrete models stands on many simplifications or the necessity of usage of some subjective assumptions ex ante, only then the estimation of the unknown parameters is even possible.

The generalization of the discrete model, that means the increasing number of particular sub-populations of the studied population, leads to the continuous form of the model. In that case the possible distribution of frailty in the population has to be defined and also the baseline mortality (it could be also formulated in a non-parametrical form). Most frequently the combination of the Gompertz function of mortality and the Gamma function of frailty is used. In the Gamma function of frailty often for simplification the average frailty at the beginning of the study (at the age or in the time $x = 0$) is considered to be one.

This method was illustrated by a simple example of a fictitious cohort; in fact a group of people deceased in 2009 was analyzed, so they were born during ca 100 past years. But it was possible to use the individual lengths of life for the calculation. This simplification corresponds in fact with the practice of the traditional period life tables' construction. It was shown that when all the assumptions hold (and if all the people who deceased during the year 2009 were really one cohort) then the average frailty of survivors decreased at lower ages for males than for females. The frailer individuals would die on average at lower age when they are males in comparison to females. Frailer females (less resistant) can survive ca 10 years longer than males with the same level of frailty. However, in both sexes only a small group of extremely resistant individuals survive to the highest ages.

The example is really only an illustration of the application to the Czech data with the aim to introduce the whole concept and its ideas and main assumptions. It is clear that for better data (cohort data and ideally cohort individual data) those introduced methods could be used better, their usage is advantageous for example in epidemiology or medicine when the surviving of patients with various risk factors is studied, etc. The individual data are often available for such studies.

For a comparison of mortality in time or space the baseline hazard should be used, (i.e. the mortality of the average individual or of an individual with the frailty equal to one), rather than the estimation of the overall mortality of the whole population because it is influenced (especially at higher ages) by the structure of surviving individuals according to their frailty. It is clear that this type of analysis will be probably more developed in the future and also new possibilities of their usage in practice will be revealed.

Whatever poet, orator, or sage may say of it,
old age is still old age

Henry Wadsworth Longfellow

Chapter 11

Summary

During the time of finishing this Thesis demography "celebrated" its 350th anniversary (February 2012). During this 3 and half centuries of development many things changed. However, what remains is the general interest in human mortality, longevity and possible limits of life span. Many demographers were dealing with these issues and many methods of mortality analysis have been developed. At the end of the 20th century and at the beginning of the 21st century this theme gained a new importance. It was caused by its possible economic or social consequences but also by many stories proving the improvements in mortality. The best known from them is probably the case of Jeanne Louise Calment, a French woman who died at the highest ever attained age.

Jeanne Louise Calment was born on February 21st, 1875 in Arles and died on August 4th, 1997 (Robine, Allard, 1999). When she died she was 122 years old. Her life became the subject of interest when she was 110 years old after her move to a nursing home ("La Masion du Lac") in 1985 (*ibid*.). Her life was studied also with the aim to evaluate the effect of life style and environment or the effect of heritability on the length of human life (Harris, 2009). Calment's life speaks more in favor of heritability because her parents and brother died at very high age (mother was 87, father 93 and brother 97 years old when they deceased; Harris, 2009, p. 84). On the other hand, some of her relatives died at young age (e.g. daughter and grandson both died at the age of 36), however, usually as a consequence of some infectious disease or accident (*ibid*.).

Changes in mortality conditions, above all of adults and the oldest-old, could have significant consequences also for the national pension systems, financial stability of life insurance institutions or social and health facilities. All those factors lead to the effort of producing accurate and detailed forecasts of future mortality development (Gavrilova, Gavrilov, 2011).

All the forecasts should be based on demographic analyses using as high-quality data as possible. There could be many methods of mortality analysis stated and many models and tools which are used in them. The aim of this Thesis was to introduce several methods of mortality analysis and describe them more in depth. All those methods showed to be connected in some way and it could be said that all of them led generally to one goal – approach to an answer to the central question in mortality analysis, whether there exist some limit of human life.

In this last chapter of the Thesis all the main ideas and conclusions of the previous parts of the text will be summarized. After this general introductory part the main theories and hypotheses connected with the initial question will be reminded. From their nature all the following methods and their partial goals could be derived. Then the introduced methodological approaches will be repeated together with the main results, their most important features and extensions. Finally, the last remarks will mention some issues omitted in this Thesis, which may be suitable for a future research.

11.1 The most important theoretical approaches

Many (if not all) the analytical methods within the mortality analysis of adults and the oldest-old population are based on some theories or assumptions about the expected or possible future development. In the background there stands the central question of the analysis, whether some limit of the life span and its improvements exists. Many theories are connected with this issue. According to their expected or proclaimed answer to the stated question, three basic groups of demographers or related scientists could be distinguished. The first of them, the optimists (visionaries) suppose that any limit of the human life does not exist or that such a limit is still growing and we are not approaching to it. Members of the second group, the empiricists, are thinking about the future outlooks through the present and past trends of the mortality development. Based on that they suppose also an increase of the life expectancy in the future (to ca 100 years during this century), but not as significant as the optimists. The last group, the pessimists (or traditionalists), suppose that the life expectancy (or the human life) cannot increase significantly in comparison with the contemporary values. Therefore they expect that the life expectancy will remain also in the future at values around ca 85 years (Manton *et al.*, 1991).

All the three groups of scientists work in their argumentations with three main hypotheses (described e.g. by Wilmoth, 1997). When the theoretical limit of life expectancy or of human life exists, then at least one of the following hypotheses would be confirmed. The first hypothesis supposes a really fixed limit of life expectancy. In fact that means that a human could live to some given age and it is not possible to survive that age for even an only day ("limited-life-span hypothesis"). This hypothesis could be taken as disproved already by this its main assumption, because it could be argued that from the statistical point of view it holds that "where it could be *n* units of something [e.g. years of age] there it could be also *n + 1* of these units"[30]. In other words that means that there is no reason why it should be possible to live to any given age which cannot be survived by only one day.

The second hypothesis ("limit-distribution hypothesis") supposes that even if the intensity of mortality is decreasing at any age, always there remains some minimal level of mortality

[30] from individual discussion with doc. F. Koschin.

which cannot be eliminated. Through these "rests" of mortality the life expectancy (or life span) cannot grow infinitely.

The third main hypothesis ("compression-rectangularization hypothesis") developed in the field of demographic analysis into an almost independent method. It studies the development of the shapes of some selected life table functions (distribution of deaths or the survival function). The main idea is based on the assumption that the survival curve approaches the rectangle as a consequence of the existence of a limit of life span. An individual part of the Thesis was devoted to the rectangularization process (Chapter 7) where it was shown that the interpretation of the results is not as straightforward as it may seem from the theory.

Naturally, there are many other theories connected with the mortality process in demography. In this Thesis mostly those somehow connected with adult and higher ages are mentioned. One group of ideas dealing with the mortality process in a theoretical way tries to explain the mortality pattern at higher ages (see e.g. Vaupel, 1997; Gavrilov, Gavrilova, 1991; Harris, 2009). Usually it is tied with the study of behavior (behavioral theories supposing altruism and support across generations) or genetic aspects (theories based on the study of DNA, hormones and cells).

11.2 Graphical methods

Graphical methods used within the Thesis and theoretically useful also in common demographic analysis were introduced in the first part of the Thesis. In the Chapter 2 the focus was aimed on those methods which allow studying of the selected demographic process in two dimensions – according to age and time. There were two different methods shown.

The first of the introduced methods was illustrated through the usage of the R software. The easy way how to construct the so-called mortality surface was presented. In this type of graphs the two dimensions (age and time) are on both the main axes and the intensity of the shown process is distinguished by different colors. Through such type of graphs the mortality development in a selected country could be studied. It allows also the comparison of the mortality development in two different populations. Within this Thesis the mortality surfaces could be found for example in the Chapter 8 where the development of the so-called senescent mortality was presented.

Another graphical methods introduced in the 2nd chapter are the 3-dimensional graphs. There are more possibilities how to prepare such type of graphs but within this work the SAS software was selected. There are more reasons for this choice. First of all, the preparation of the graph is easy and quick in SAS. Secondly, the output could be simply adjusted to actual needs. Third, the graphs exported to the html-format may be used for attractive presentation when they could be rotated or zoomed. The 3-dimensional graphs were used also later in the work, for example in the Chapter 4, 7 or 8. Many other results could be presented in this manner.

Thanks to the introduced types of graphs also the cohort effects could be studied without the need of using some more sophisticated methods. It could be said, that the advanced or more complicated analytical methods could be selected (and then applied) after the first graphical analysis of the data and its development.

11.3 Methodological approaches to mortality

When we focus solely on the mortality process, then the unexpectedly rapid growth of numbers as well as of proportions of survivors to higher ages occurs in a population of many demographically developed countries, above all numbers and proportions of the oldest-old people. This growth was a consequence of improving mortality conditions at all ages but especially at the higher and highest ages. Other important fact is connected with the improving availability of reliable and more detailed demographic data and the latest development of new statistical software and hardware or computation methods.

Thanks to all those factors and reasons still more demographers deal with the issue of mortality development and mortality changes. Also a growing interest in proper methods or elaboration of the existing traditional ones could be traced in demography as a whole. Among others, it is a logical consequence of greater possibilities of the demographic analysis (given by the mentioned more reliable data and more sophisticated software). All these aspects stand also behind the existence of this Doctoral Thesis focused just on the more detailed elaboration of the traditional methods and evaluation or introduction of other, more specific, ones.

11.3.1 Life tables

The first part of the Thesis was devoted mainly to the theme of life tables which could be taken as a standard and probably the most traditional tool of demographic analysis. The issue was started by the detailed introduction to the current construction method of the official life tables in the Czech Republic (produced by the Czech Statistical Office) which is based on the application of the Gompertz-Makeham method. This method is common in more, above all Eastern European, countries (e.g. Slovakia or Estonia; European Commission, 2003) but it is used also in several analytical studies or for the description of the mortality pattern in general (as mention e.g. Gavrilov, Gavrilova, 1991; Gavrilova, Gavrilov, 2011). It was shown that the King-Hardy method of parameter estimation has several weaknesses (e.g. its sensitivity to the subjective choice of the input parameters). It was the motive for searching of several alternatives how to deal with the task (usually taken as one of the simplest ones in demography) of life table construction.

It was mentioned that there are more alternatives of a change in life table construction (for example the ending of the table). However, the method of smoothing and extrapolation of the intensity of mortality is, without any doubt, one of the most important ones. The question is not only about the selection of a proper method but also about finding any relevant and accurate approach to its parameter estimation (as we do not consider the non- or semi-parametric methods within this work). Several, the most important ones, mortality laws (or, in general, functions describing the mortality pattern with age) are briefly described in the text. But in more detail they are introduced in an article published within the preparation of this Thesis (Burcin *et al.*, 2010) or in other relevant literature (Boleslawski, Tabeau, 2001; Thatcher *et al.*, 1998; Caselli, 2006). Within this work, there more attention was paid to the method of the parameter estimation. Because the proposed method of non-linear weighted least squares is relatively complicated to be calculated without any special statistical tools or applications a simple programmed application for the SAS software was prepared and attached to this Thesis in an electronic form. Also not a very advanced user of the SAS software can use this code for producing the estimates of the unknown parameters or directly the smoothed values of the intensity of mortality based upon 6 selected (probably the most frequently used) mortality laws. This macro was also used in other parts of the Thesis where the smoothed values of mortality rates were necessary.

The programmed code builds upon several particular and rather individual macros solving each model separately. In this way the code could be enriched in the future easily. The aim of this code is not to compete with other similar applications or software (e.g. the independent software application "DeRaS" was developed while working on this issue[31] offering not only smoothing of the mortality rates or probabilities of dying but also the whole life table construction) but rather to offer a simplified alternative which could be helpful in such specific situations where only the estimates of the parameters or the smoothed intensities of mortality are needed.

Other issues related to the life table construction and solved within this Thesis are connected with the accessibility of more detailed (often individual) mortality data. This availability enables to the demographer to consider more approaches to the mortality rates calculation. In Chapter 6 the methodology of the age-specific mortality rates calculation was described and showed in detail. According to the description all three possible methods (based on the 1st, 2nd, and also 3rd classes of demographic events) were used in the analysis.

Based on the comparison of the results obtained from the method of individual life durations and from the traditional method[32] the more detailed calculation was preferred in the following part of the text. Individual life durations were also used in the calculation of the exposure time of the deceased persons for the construction of the life tables according to age, sex and education attainment. Through these tables it was verified that there is a clear relationship between education and mortality levels in the population. However, from the methodological point of view the differences according to the used method of mortality rates construction were proved.

Some interesting conclusions of Rychtaříková (2010) were also verified in this way. In her research she showed an anomaly indicating that for females in the Czech Republic there is almost no difference in the probability of dying between the group of females with no or only primary education and females with lower secondary education. For some ages the relation is even reversal to the expected one. It confirmed the fact that mortality level is highly connected with the prevailing life style or type of work. It might be an interesting research question to study this relation also in other countries.

11.3.2 Rectangularization process, mortality shifting

The second part of the Thesis deals with several topics of mortality analysis, which might seem to be almost untied on the first sight. However, the opposite is true. The first issue solved in this

[31] Burcin *et al.*, 2011a

[32] The "traditional method" uses the average of the numbers of survivors at the beginning and at the end of the year as the estimate of the exposure time.

part of the text is closely related to the previous theme, the life tables. In the 7th chapter several indicators of the process of rectangularization of the survival curve (or compression of mortality) are introduced. Through these indicators not only the process is illustrated in the Eastern as well as in Western or Northern European countries but also the main idea of this concept is introduced. As was explained above the whole idea stands on the assumption that there exists some fixed limit of life span. Then as a consequence of improvements in mortality the shape of the survival function of the studied population approaches to a rectangle.

Many indicators suitable for the analysis of the rectangularization process were introduced there. Through these indicators it was illustrated that their development shows some signs of stagnation during the latest years, above all in the most developed low-mortality countries (like Sweden). As was derived in this work, in the countries (or in such situations) where the mortality conditions get worse this process could be called “derectangularization”, and only in those countries (or situations) where the positive trend continues the stagnation of the traditional indicators is more connected with the so-called “shifting”. That means that for example Sweden (especially females) could be taken as a representative of such a country where the mortality shifting could be expected. On the other hand, in Russia and some other post-communist countries the process of derectangularization could be seen, at least in some historical periods. Usually this process in those countries is connected also with worsening of the economic or social situation.

The process of mortality shifting is more deeply studied in the Chapter 8. Partly the issue was solved in a formal way when the supposed relations were derived only theoretically. Then several selected indicators used in the analysis of rectangularization were chosen for the application in this part with the aim to find some general impulses leading to the process of mortality shifting. Those indicators were the interquartile range, modal age at death and the standard deviation of ages at death above the mode. Then two model mortality data sets were generated – following the exponential and the logistic pattern. For these model data the mortality improvements were simulated – as the absolute or relative decreases of the age-specific mortality rates. Through the application of the selected indicators it was shown that the “pure” mortality shifting would appear for exponentially increasing mortality where the constant relative decrease of mortality rates occurs.

The process of mortality shifting could be (with some simplification) taken as the currently most favorable mortality development, where the deaths are postponed to higher ages. Then the results of this development are the horizontal shift of the life table functions (the survival function or the density of deaths were used within this Thesis) to higher ages, higher concentration of survivors at higher ages and the increase of the modal age at death. This process could be also understand as a natural follow-up of the rectangularization process when the deaths stop to be more concentrated around the mode and the whole curve starts to be shifted to the right.

11.3.3 Tempo effects in mortality analysis

The analysis of so-called tempo effects, involved in the 9th chapter, could be understood almost as a response to the analysis of the previous process (mortality shifting). In fact it is a study of the effects of the modal (and mean) age at death increase to the standard and traditional measures of the mortality process. Within this part, there the existing ways of adjusting of the traditional measures were introduced and applied to the empirical data. However, it showed out that one important assumption of this ideological concept is not fulfilled for many historical but also contemporary populations (the assumption of horizontal shift of the mortality curve without any change of its shape). This fact can influence significantly the study of the tempo effects.

When it could be assumed that the mentioned condition does not hold, then some alternative methods of the tempo adjustment can be used. One of them was developed by Kohler and Philipov (2001). This method was originally proposed for the process of fertility. Within this Thesis this methodology was applied to the mortality data. Then also a simplification of it was proposed because in the field of mortality some calculations could be omitted. For example in mortality analysis we can use the knowledge of the supposed value of the so-called total mortality rate. In the case of fertility the total number of live births could differ among females (and it differs significantly). However, in the case of mortality each person could (and have to) die only once.

The results confirm that the tempo-adjusted measures are lower than the un-adjusted ones. Because these methods are often highly data-demanding they could be fully applied to data from mainly the Western or Northern European countries where long time series of reliable data are available.

11.3.4 Unobserved heterogeneity in data and frailty models

The last chapter of the methodological part of the Thesis was devoted to the introduction of the frailty models. These models alone would be enough for writing an individual Thesis (as did e.g. Wienke, 2011). This issue represents an alternative approach to the study of mortality patterns based on the assumption of the existence of the so-called unobserved heterogeneity. In fact that means that each person (or group of persons) in the population could be characterized by a concrete level of a specific factor. It is the main idea of the survival analysis in demography. The survival analysis studies the time durations from some initial event to some other defined event according to the various levels of the selected observable factors. Frailty models are only a special type of the proportional hazards models in the survival analysis where instead of the observable factors at least one unobservable factor is considered – the frailty. If everyone in the population could be characterized by his or her specific level of frailty then the population could be taken as a heterogeneous one.

In the Chapter 10 the discrete as well as the continuous frailty models were described. Surprisingly the continuous models are not so difficult for calculation (not so many parameters have to be estimated) in comparison to the discrete ones. Frailty models enable to study the changes of the population structure according to frailty. Also the “average person” (again according to frailty) could be described by the baseline intensity of mortality. Finally it was concluded that mortality pattern of this “average person” (i.e. the baseline hazard) could be used for comparison of mortality among different populations. The baseline mortality hazard (or baseline intensity of mortality) is not affected by the effect of the composition of population according to frailty. On the other side the traditional measures representing the population-based intensity are influenced by the unobserved heterogeneity of the total population.

Frailty models are primarily derived for cohort data. However, in the example it was illustrated that also the results obtained for the period data (when these are reliable and detailed enough) could reveal some interesting aspects influencing the overall mortality development, e.g. the composition of the total population according to the levels of frailty of the survivors. It was shown that the average level of frailty in the population decreases more rapidly for males in comparison to females. It is the consequence of generally higher level of mortality of males. Because of that females could live some 10 years longer in comparison to males with the same level of frailty. For both sexes then it was confirmed that at the extreme ages the concrete level of frailty doesn't play a significant role because only a small group of the most robust individuals survive to that age.

11.4 The final remarks

This Thesis is one from, in the Czech demography still not very common, rather methodological works. Its aim was not to compete with the advanced and developed foreign methodological works or studies usually focused more in depth on a narrow range of themes. Instead of that this Thesis brings to the (mainly Czech or Eastern European) reader the possibility to learn some maybe new and interesting approaches or to develop in more detail the traditional and used ones. All the methods introduced in this work were applied also to the real data so that the results could be seen together with its brief interpretation.

All of the introduced methods in the Thesis have one aspect in common – they are at least trying to find any way how to prove the existence or nonexistence of the theoretical limit of life span. Therefore all the involved methods study various aspects of the same mortality process and all represent particular ways how to study the overall mortality development. Then all the methods are in fact different answers to the initial question about the limit of human life span.

As was said already at the very beginning of the Thesis, the aim of it was not to form a textbook containing all the usable methods related to the mortality analysis. Moreover, it is not possible to cover all the approaches. This Thesis was mainly focused on methods related to the life tables or to the study of the period data. However, at the end it was clear that the frailty models as well as the analysis of tempo-effect are more or less connected with the cohort data. Study of cohort effects in period data or the whole range of methods proposed for cohort data was however omitted in this work. That means, that any future continuation of the issue of mortality analysis would be (with respect to still better quality and detail of demographic data) oriented more strongly to the cohort analysis.

Another aspect which was mentioned only in some parts of the Thesis was the availability of detailed, often individual, data. Within the work in some cases there the individual data were used – e.g. for the calculation of the time exposure in age-specific mortality rates and in the frailty models. Of course, the potential of the individual data analysis is much wider. It could be taken as a second source of theoretical future extension of the topic of mortality analysis – to focus on methods suitable to individual or detailed data. Those could be for example several types of survival analysis, regressions expressing the influence of various factors, or detailed analyses based on the study of developmental trends across age, time and cohorts.

Hopefully, now it is clear that although this Thesis was divided into several parts which may seem to be independent on each other, this view would be too simplified. In fact, all the parts of the Thesis are connected and tied together, as it reflects the width and interconnection of the possible points of view applicable to the mortality process and mortality analysis. Several of the introduced methods lead to the similar results or enable to study the same developmental trends taken only from various perspectives. But sometimes it might be useful to use wider range of proper methods because, as was shown, one supports the other or it could be used for evaluation of the assumptions, for placing the studied issue into wider relations, or for finding some interesting details in the studied phenomenon.

It could be concluded that many (if not all) of the introduced approaches and concepts were derived with the immodest (nevertheless natural) question whether there exist some limit of the human life. Although all the methods were applied to model or real data, and although all the methods were studied in more detail, none of them could verify or even prove the existence, but also the non-existence, of limits of human life. So this question remains open and probably it will remain also in the future.

In the introductory part of this Thesis, there were several partial aims stated, these were solved in particular chapters of the work. But it was also mentioned there, that the main task is to ask the questions related to the overall issue of the Thesis and to try to find or formulate some possible answers. Then, the main conclusion in this aspect could be the fact, that the more we know, the more questions we have. Such an effect opens not only the human mind but also, more concretely, future possibilities of the demographic research. The author would be happy to find out that at least one future reader of this Thesis was inspired by it for his or her own research. It is possible to build upon the partial conclusions introduced in the text of the work or to follow his or her own questions raised from the issues involved in the previous chapters.

Considering the initial question whether there exist some limit of the human life – maybe one day we will find the answer… and maybe we will learn that the knowledge of such a potential limit is for the quality of our lives not important at all…

REFERENCES

AALEN, O.O.; BORGAN, Ø.; GJESSING, H.K. 2010. *Survival and event history analysis : A process point of view*. New York : Springer, 2010. ISBN 978-1-4419-1909-0.

ANDREEV, E.M.; SHKOLNIKOV, V.M.; BEGUN, A.Z. 2002. Algorithm for decomposition of differences between aggregate demographic measures and its application to life expectancies, healthy life expectancies, parity-progression ratios and total fertility rates. *Demographic Research.* 2002, Vol. 7, No. 14, pp. 499–522. <http://www.demographic-research.org/Volumes/Vol7/14/7-14.pdf>.

BARBI, E.; CASELLI, G.; YASHIN, A. 2004. Age and time patterns of mortality by cause in Italy : A mortality surface approach. *Proceedings of the XLII Scientific Meeting of the Italian Society of Statistics*, Bari 9.–11. June, 2004.

BARBI, E.; BONGAARTS, J.; VAUPEL, J.W. (eds.). 2008. How long do we live? : Demographic models and reflections on tempo effects. Berlin : Springer, 2008, *Demographic Research Monographs 5*. ISBN 978-3-540-78519-4. <http://www.demogr.mpg.de/en/projects_publications/publications_1904/monographs/how_long_do_we_live_demographic_models_and_reflections_on_tempo_effects_3080.htm >.

BBC News, 2004. *Humans "will live to age of 150"*. BBC News, 22. 10. 2004. <http://news.bbc.co.uk/2/hi/science/nature/3761310.stm>.

BEARD, E.R. 1959. Note on some mathematical mortality models. In: WOLSTENHOLME, G.E.W.; O'CONNER, M. (eds.). *The Lifespan of Animals*. Boston : Ciba Foundation Colloquium on Aging, 1959, pp. 302–311.

BOLESLAWSKI, L.; TABEAU, E. 2001. Comparing theoretical age patterns of mortality beyond the age of 80. In: TABEAU, E.; VAN DEN BERG JETHS, A.; HEATHCOTE, CH. (eds.). *Forecasting Mortality in Developed Countries : Insights from a Statistical, Demographic and Epidemiological Perspective.* Springer Netherlands, 2001, pp. 127–155. ISBN 978-0-7923-6833-5. <http://www.springerlink.com/content/vv65332142w06u2h/>.

BONGAARTS, J. 2005. Long-range trends in adult mortality : Models and projection methods. *Demography*. February 2005, Vol. 42, No. 1, pp. 23–49.

BONGAARTS, J. 2009. Trends in senescent life expectancy. *Population Studies*. 2009, Vol. 63, No. 3, pp. 203–213.

BONGAARTS, J. 2012. Individual e-mail consultation. [2012-03-20]

BONGAARTS, J.; FEENEY, G. 1998. On the quantum and tempo of fertility. *Population and Development Review*. 1998, Vol. 24, No. 2, pp. 271–291. <http://www.popcouncil.org/pdfs/councilarticles/pdr/PDR242Bongaarts.pdf>.

BONGAARTS, J.; FEENEY, G. 2002. How Long Do We Live? *Population and Development Review*. 2002, Vol. 28, No. 1, pp. 13–29.

BONGAARTS, J.; FEENEY, G. 2003. Estimating mean lifetime. Population Council, Policy Research Division, 2003, *Working Paper. No. 179*, 25 pp. <http://www.popcouncil.org/pdfs/wp/179.pdf>.

BONGAARTS, J.; FEENEY, G. 2006. The quantum and tempo of life-cycle events. *Vienna Yearbook of Population Research.* 2006, pp. 115–151.

BONGAARTS, J.; FEENEY, G. 2008. The quantum and tempo of life-cycle events. In: BARBI, E.; BONGAARTS, J.; VAUPEL, J.W. (eds.). How long do we live? : Demographic models and reflections on tempo effects. Berlin : Springer, 2008, *Demographic Research Monographs 5*, pp. 29–65. <http://www.demogr.mpg.de/books/drm/005/3.pdf>.

BURCIN, B.; KUČERA, T.; DRBOHLAV, D. 2003. *Perspektivy populačního vývoje České republiky na období 2003–2065.* Praha : DemoArt, 2003. 29 pp. ISBN 80-86746-01-1.

BURCIN, B.; ŠÍDLO, L.; TESÁRKOVÁ, K. 2009. *Rectangularization process of the survival curve in selected European countries.* Paper presented at the V. International antropological congress of Aleš Hrdlička “Quo vadis homo ...societas humana?”, Praha, Humpolec, 2.–5. 9. 2009.

BURCIN, B.; TESÁRKOVÁ, K.; ŠÍDLO, L. 2010. Nejpoužívanější metody vyrovnávání a extrapolace křivky úmrtnosti a jejich aplikace na českou populaci. *Demografie*. 2010, Vol. 52, No. 2, pp. 77–89. ISSN 0011-8265.

BURCIN, B.; HULÍKOVÁ TESÁRKOVÁ, K. 2011. *Úmrtnost obyvatelstva České a Slovenské republiky, návrh konstrukce nových úmrtnostních tabulek.* Paper presented at the XLI. Conference of the Czech Demographic Society “Česko a Slovensko ve společném státě a samostatně, podobnosti a odlišnosti”, 26.–27. 5. 2010, Olomouc, Faculty of Science, Palacký University, Olomouc, Czech Republic.

BURCIN, B.; HULÍKOVÁ TESÁRKOVÁ, K.; KOMÁNEK, D.; KUČERA, T. 2011a. *DeRaS: software tool for modelling mortality intensities and life table construction*. Charles University in Prague, 2011. <http://www.deras.natur.cuni.cz/>.

BURCIN, B.; HULÍKOVÁ TESÁRKOVÁ, K.; KOMÁNEK, D. 2011b. *DeRaS – a new SW application for the life table construction*. Poster presented at the International Workshop “Health, Morbidity and Mortality: the State of the Art”, 1.–3. 9. 2011, Vienna Institute of Demography, Vienna, Austria.

CALOT, G.; FRANCO, A. 2002. The construction of life tables. In: WUNSCH, G.; MOUCHART, M.; DUCHÊNE, J. (eds.). 2002. *The Life Table : Modelling Survival and Death*. Kluwer Acadeic Publisher, Dordrecht, The Netherlands, 2002, European studies of population, European association for population studies, pp. 33–78. ISBN 1-4020-0638-1.

CAMPION, E.W. 1994. Editorial. *The New England Journal of Medicine.* 1994, Vol. 330, No. 25, pp. 1819–1820.

CANUDAS-ROMO, V. 2008. The modal age at death and the shifting mortality hypothesis. *Demographic Research.* 2008, Vol. 19, Article 30, pp. 1179–1204. <http://www.demographic-research.org/volumes/vol19/30/19-30.pdf>.

CAREY, J.R. *et al.* 1995. A male-female longevity paradox in medfly cohorts. *Journal of Animal Ecology*. 1995, Vol. 64, No. 1, pp. 107–116.

CAREY, J.R.; GRUENFELDER, C. 1997. Population biology of the elderly. In: WACHTER, K.W.; FINCH, C.E. (eds.). *Between Zeus and the Salmon : Biodemography of Longevity*. Washington, D.C. : National Academy Press, 1997, pp. 127–160.

CASELLI, G. 2006. Mortality forecasts: Hypotheses and methods. In: CASELLI, G.; VALLIN, J.; WUNSCH, G. (eds.). *Demography : Analysis and Synthesis*. Academic Press, 2006, Vol. 3, pp. 177–188, ISBN 012765660X.

COELHO, E.; MAGALHÃES, M.G.; BRAVO, J.M. 2007. Mortality and longevity projections for the oldest-old in Portugal. *Proceedings of the Joint Eurostat/UNECE (United Nations Statistical Commission and Statistical Office of the Economic Commission for Europe) Work Session on Demographic Projections.* Bucharest, Romania, 2007, pp. 117–132. <http://epc2008.princeton.edu/download.aspx?submissionId=80105>.

COX, D.R. 1972. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*. Series B, 1972, No. 34, pp. 187–220.

CURTSINGER, J.W. 1995. Density, mortality, and the narrow view. *Genetica.* 1995, Vol. 96, No. 3, pp. 187–189. <https://springerlink3.metapress.com/content/j47021011w80h845/resource-secured/?target=fulltext.pdf&sid=4fitr2a2e04gwakkaigmhykq&sh =www.springerlink.com>.

Czech Statistical Office. 2009a. *Úmrtnostní tabulky za ČR, oblasti a kraje 2008.* 2009 [cit. 2010-04-25.]. <http://www.czso.cz/csu/2009edicniplan.nsf/p/4002-09>.

Czech Statistical Office. 2009b. *Úmrtnostní tabulky za ČR, oblasti a kraje 2008 : Metodické poznámky*. 2009 [cit. 2009-12-01]. <http://czso.cz/csu/2009edicniplan.nsf/t/BD00408CCA/$File/400209m.pdf>.

Czech Statistical Office. 2010–2012. *Demografie – pramenná díla*. Database of data from Pohyb obyvatelstva v České republice. Accessed in 2010–2012 at <http://www.czso.cz/csu/redakce.nsf/i/casova_rada_demografie>.

Czech Statistical Office. 2011a. *Pohyb obyvatel za roky 2009, 2010.* Unpublished individual data, 2011.

Czech Statistical Office. 2011b. *Demografická ročenka ČR 2010.* <http://www.czso.cz/csu/2011edicniplan.nsf/p/4019-11>.

DE GREY, A. 2004. *We will be able to live to 1,000*. BBC News, 3. 12. 2004. <http://news.bbc.co.uk/2/hi/uk/4003063.stm>.

European Commission, 2003: *Demographic statistics: definitions and methods of collection in 31 European countries*. Population and social conditions 3/2003/E/n° 25, Luxembourg: Office for Official Publications of the European Communities, 2003, 117 pp.

FEENEY, G. 2006. Increments to life and mortality tempo. *Demographic Research.* 2006, Vol. 14, Article 2, pp. 27–46. <http://www.demographic-research.org/volumes/vol14/2/14-2.pdf>.

FIALA, T. 2005. *Výpočty aktuárské demografie v tabulkovém procesoru*. Praha : Oeconomica, 2005, 177 pp. ISBN 80-245-0821-4.

FIALA, T.; LANGHAMROVÁ, J.; PRŮŠA, L. 2011. Projection of the human capital of the Czech Republic and its regions to 2050. *Demografie*. 2011, Vol. 53, No. 7, pp. 304–320.

FRIES, J.F. 1980. Aging, natural death, and the compression of morbidity. *The New England Journal of Medicine.* 1980, pp. 130–135. [cit. 2009-02-16]. <http://www.scielosp.org/pdf/bwho/v80n3/v80n3a12.pdf>.

GAVRILOV, L.A.; GAVRILOVA, N.S. 1979. Historical perspective on kinetics of human mortality. *Proceedings of the USSR Academy of Sciences* [Doklady Akademii Nauk SSSR]. 1979, Vol. 245, No. 4, pp. 1017–1020.

GAVRILOV, L.A.; GAVRILOVA, N.S. 1991. *The biology of life span : A quantitative approach.* Harwood Academic Publishers GmbH, Chur, Switzerland, 1991, 385 pp.

GAVRILOVA, N.S.; GAVRILOV, L.A. 2011. Ageing and longevity: Mortality laws and mortality forecasts for ageing populations [In Czech: Stárnutí a dlouhověkost: Zákony a prognózy úmrtnosti pro stárnoucí populace]. *Demografie*. 2011, Vol. 53, No. 2, pp. 109–128. <http://longevity-science.org/pdf/Demografie-English-2011.pdf>. [online in English]

GAVRILOVA, N.S.; SEMANYOVA, V.G.; EVDOKUSHKINA, G.N. 2002. Mortality crisis in Russia : New health threats. Unraveling the secrets of human longevity. *Russian Studies : Russian Health, Mortality and Lifespan* [online]. Chicago : 2002 [cit. 2009-08-02]. <http://longevity-science.org/PAA-2002-Russia.pdf>.

GERYLOVÁ, A.; HOLČÍK, J. 1988*. Základy statistické metodologie.* Brno : Univerzita Jana Evangelisty Purkyně, Fakulta lékařská, 1988, 128 pp.

GOLDSTEIN, J.R. 2006. Found in translation? A cohort perspective on tempo-adjusted life expectancy. *Demographic Research.* 2006, Vol. 14, No. 5, pp. 71–84.

GOMPERTZ, B. 1825. On the nature of the function expressive of the law of human mortality etc. *Philosophical Transactions of the Royal Society London*, 1825, 115, pp. 513–585.

HARRIS, T.F. 2009. *Living to 100 and Beyond*. ACTEX Publications, Inc. 2009, 240 pp. ISBN 978-1-56698-699-1.

HARTMANNOVÁ, H.; FESENKO, P. 1973. Úmrtnostní tabulky 1970. *Demografie*. 1973, Vol. 15, No. 3, pp. 235–241.

HAYFLICK, L. 1965. The limited in vitro lifetime of human diploid cell strains. *Experimental Cell Research*. March 1965, Vol. 37, No. 3, pp. 614–636. <http://www.sciencedirect.com/science/article/pii/0014482765902119>.

HAZAFYOVÁ, M. 2009. *Fenomén aktivního stáří v Pardubicích*. Pardubice, 2009. Bakalářská práce (Bc). Univerzita Pardubice, Fakulta filozofická.

HOLMAN, R. 2002. *Ekonomie*. Praha : C.H. Beck, 2002, 714 pp. ISBN 80-7179-681-6.

HOUGAARD, P. 1999. Fundamentals of survival data. *Biometrics*. March 1999, No. 55, pp. 13–22.

HULÍK, V.; FIALA, T. 2011. *Projekce vzdělanostní struktury obyvatelstva do roku 2050.* Unpublished data from the forecast prepared in the research project No. 2D06026 “Reprodukce lidského kapitálu”. Praha : Ústav pro informace ve vzdělávání, 2011.

HULÍKOVÁ TESÁRKOVÁ, K. 2011. *Možnosti zobrazení vývoje úmrtnosti v České republice a na Slovensku ve statistickém software SAS a R*. Poster presented at the XLI. Conference of the Czech Demographic Society "Česko a Slovensko ve společném státě a samostatně, podobnosti a odlišnosti", 26.–27. 5. 2010, Olomouc, Faculty of Science, Palacký University, Olomouc, Czech Republic.

HULÍKOVÁ TESÁRKOVÁ, K. 2012: Modely křehkosti a jejich aplikace v demografii. *Demografie*. 2012, Vol. 54, No. 1, pp. 25–36.

Human Mortality Database. 2010. Accessed in August 2010 at <http://www.mortality.org>.

HUSTEAD, E.C. 2005. *Ending the Mortality Table.* Paper presented at the Living to 100 and Beyond Symposium, 2005,Orlando, Fla : Society of Actuaries, January 12–14, 2005.

CHEUNG, S.L.K.; ROBINE, J.M.; TU, E.J.; CASELLI, G. 2005. Three dimensions of the survival curve: Horizontalization, verticalization, and longevity extension. *Demography*. 2005, Vol. 42, No. 2, pp. 243–258.

CHOU, K.; CHI, I. 2002. Successful aging among the young-old, old-old, and oldest-old Chinese. *The International Journal of Aging & Human Development.* 2002, Vol. 54, No. 1, pp. 1–14.

JOHNSON, T.E.; SHOOK, D.R. 1997. Identification and mapping of genes determining longevity. In: WACHTER, K.W.; FINCH, C.E. (eds.). *Between Zeus and the Salmon : Biodemography of Longevity*. Washington, D.C. : National Academy Press, 1997, pp. 108–126. ISBN 0-309-05787-6.

JŮZOVÁ, J. 2006. *Možnosti přípravy na stáří*. Brno, 2006. Diplomová práce (Mgr). Masarykova univerzita, Filozofická fakulta. Ústav pedagogických věd.

KAFKOVÁ, M. 2006. *Postoje populace ke stárnutí v evropských zemích*. Brno, 2006. Diplomová práce (Mgr). Masarykova univerzita. Fakulta sociálních studií. Katedra sociologie.

KANNISTO, V. 1996: *The Advancing Frontier of Survival: Life Tables for Old Age*. Odense University Press : Odense Monographs on Population Aging, 1996, Vol. 3. ISBN 87-7838-185-1. <http://www.demogr.mpg.de/Papers/Books/Monograph3/start.htm>.

KANNISTO, V. 2000. Measuring the compression of mortality. *Demographic Research.* 2000, Vol. 3, Article 6. <http://www.demographic-research.org/volumes/vol3/6/3-6.pdf>.

KANNISTO, V. 2001. Mode and dispersion of the length of life. *Population: an English Selection.* 2001, Vol. 13, No. 1, pp. 159–171. <http://www.persee.fr/web/revues/home/prescript/article/pop_0032-4663_2001_hos_13_1_7235>.

KANNISTO, V.; LAURITSEN, J.; THATCHER, A.R.; VAUPEL, J.W. 1994. Reductions in mortality at advanced ages : Several decades of evidence from 27 countries. *Population and Development Review.* 1994, Vol. 20, No. 4, pp. 793–810. ISSN 0098-7921.

KOCOVÁ, M. 2009. *Ruská úmrtnostní krize a její důsledky pro budoucí vývoj společnosti.* Praha, 2009. Bakalářská práce (Bc). Univerzita Karlova v Praze. Přírodovědecká fakulta. Katedra demografie a geodemografie.

KOHLER, H.P.; PHILIPOV, D. 2001. Variance effects in the Bongaarts-Feeney formula. *Demography*, February 2001, Vol. 38, No. 1, pp. 1–16.

KOSCHIN, F. 1981. Gompertz-Makehamův zákon aneb o přesnosti. *Demografie*. 1981, Vol. 23, No. 3, pp. 207–220.

KOSCHIN, F. 1989. Mortality and age structure at old ages. *Acta Demographica*. 1989, Vol. IX, No. 3, Ageing of population in developed countries : Causes – consequences – policies, International Population Conference, Czechoslovak Demographic Society of the Czechoslovak Academy of Sciences, Prague, Czechoslovakia, 4.–7. 7. 1989, pp. 270–280.

KOSCHIN, F. 1999. Jak vysoká je intenzita úmrtnosti na konci lidského života? *Demografie*. 1999, Vol. 41, No. 2, pp. 105–119.

KOSCHIN, F. 2002. *Aktuárská demografie.* Praha : Nakladatelství Oeconomica, 2002, 100 pp. ISBN 80-245-0403-0.

KOSCHIN, F.; FIALA, T.; LANGHAMROVÁ, J.; ROUBÍČEK, V. 1998. *Úmrtnost v českých zemích v devadesátých letech.* Praha : VŠE, 1998. 68 pp. ISBN 80-7079-574-3.

KOUDELKA, P.; LUSTIGOVÁ, M. 2010. Užití víceúrovňových a víceprocesových modelů v demografii. *Demografie*. 2010, Vol. 52, No. 1, pp. 235–248.

LUY, M. 2010a. Paradigm shift in demography? Editorial to the special issue “Tempo effects in demographic period indicators”. *Comparative Population Studies*, 2010, Vol. 35, No. 3, pp. 409–414.

LUY, M. 2010b. Tempo effects and their relevance in demographic analysis. *Comparative Population Studies*. 2010, Vol. 35, No. 3, pp. 415–446.

MAKEHAM, W.M. 1860. On the law of mortality and the construction of annuity tables. *Journal of the Institute of Actuaries and Assurance Magazine.* 1860, No. 8, pp. 301–310.

MANTON, K.G.; STALLARD, E.; TOLLEY, H.D. 1991. Limits to human life expectancy : Evidence, prospects, and implications. *Population and Development Review.* 1991, Vol. 17, No. 4, pp. 603–637.

NAGNUR, D. 1986. Rectangularization of the survival curve and entropy : The Canadian experience, 1921–1981. *Canadian Studies in Population.* 1986, Vol. 13, No. 1, pp. 83–102.

OLSHANSKY, S.J. 2004. *Don’t fall for the cult of immortality*. BBC News, 3. 12. 2004. <http://news.bbc.co.uk/2/hi/uk_news/4059549.stm>.

PASTOR, K. 2007. A cubic model for graduation of life tables. *Journal of the Applied Mathematics, Statistics and Informatics.* 2007, Vol. 3, No. 2., pp. 225–236. <http://fpv.ucm.sk/jamsi/docs/v03n02_12_2007/v03n02_08_PASTOR.pdf>.

PAVLÍK, Z.; RYCHTAŘÍKOVÁ, J.; ŠUBRTOVÁ, A. 1986. *Základy demografie.* Praha : Academia, 1986. 736 pp.

PECKA, J. 1989. Příspěvek k problematice výpočtu československých úmrtnostních tabulek. *Demografie*. 1989, Vol. 31, No. 3, pp. 229–238.

PRESTON, S.H.; HEUVELINE, P.; GUILLOT, M. 2001. *Demography: Measuring and Modeling Population Processes*. Blackwell Publishing, 291 pp. ISBN 978-1-55786-214-3.

R Development Core Team. 2009. *R : A language and environment for statistical computing*. R Foundation for statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

ROBINE, J.M.; ALLARD, M. 1999. Jeanne Calment : Validation of the duration of her life. In: JEUNE, B.; VAUPEL, J. W. (eds.). *Validation of Exceptional Longevity*. Odense University Press : Odense Monographs on Population Aging, 1999, Vol. 6. ISBN 87-7838-466-4. <http://www.demogr.mpg.de/books/odense/6/09.htm>.

RODRÍGUEZ, G. 2008. Demographic translation and tempo effects : An accelerated failure time perspective. In: BARBI, E.; BONGAARTS, J.; VAUPEL, J.W. (eds.). 2008. How long do we live? : Demographic models and reflections on tempo effects. Berlin : Springer, 2008, *Demographic Research Monographs 5*, pp. 69–92. ISBN 978-3-540-78519-4. <http://www.demogr.mpg.de/books/drm/005/4.pdf>.

ROLI, G. 2008. *An adaptive procedure for estimating and comparing the old-age mortality in a long historical perspective: Emilia-Romagna, 1871–2001*. Bologna : Dipartimento di Scienze Statistiche “Paolo Fortunati”, Alma Mater Studiorum Universita di Bologna, 2008.

ROSE, M.R. 2007. New biological books. *The Quarterly Review of Biology*. 2007, Vol. 82, No. 4, pp. 395–400. <http://www.journals.uchicago.edu/doi/pdf/ 10.1086/522812?cookieSet=1>.

ROUBÍČEK, V. 1996. *Základní problémy obecné a ekonomické demografie.* Praha : Vysoká škola ekonomická, 1996, 271 pp. ISBN 80-7079-188-8.

RYDER, N.B. 1956. Problems of trend determination during a translation in fertility. *Milbank Memorial Fund Quarterly*. 1956, Vol. 34, No. 1, pp. 5–21.

RYCHTAŘÍKOVÁ, J. 2008. Nové metody demografické analýzy. *Demografie.* 2010, Vol. 50, No. 4, pp. 250–258.

RYCHTAŘÍKOVÁ, J. 2010. *Does education impact health outcomes in the Czech Republic as expected?* Paper presented in Max Planc Institute for Demographic Research, Rostock, Germany. 18.–19. 3. 2010.

SAS Institute Inc. 2009. SAS OnlineDoc® 9.2. Cary, NC: SAS Institute Inc.

SHKOLNIKOV, V.M.; ANDREEV, E.M.; LEON, D.A.; McKEE, M.; MESLÉ, F.; VALLIN, J. 2004. Mortality reversal in Russia : The story so far. *Hygiea Internationalis*. 2004, Vol. 4, No. 4, pp. 29–80. <http://www.demogr.mpg.de/publications/files/1836_1119978429 _1_HygInt-4.2004.4.pdf>.

SHKOLNIKOV, V.M.; ANDREEV, E.M. 2010. *Age-decomposition of a difference between two populations for any life-table quantity in Excel*. MPIDR technical report 2010-002, Max Planck Institute for Demographic Research, 2010. <http://www.demogr.mpg.de/en/ projects_publications/publications_1904/mpidr_technical_reports/>.

SMITH, D.P. 1992. *Formal Demography*. Plenum Press, New York : The Plenum Series on Demographic Methods and Population Analysis, 1992, 329 pp. ISBN 0-306-43869-0.

STEINSALTZ, D.R.; WACHTER, K.W. 2006. Understanding mortality rate deceleration and heterogeneity. *Mathematical Population Studies*, 2006, No. 13, pp. 19–37.

ŠÍDLO, L.; TESÁRKOVÁ, K. 2009. *The process of rectangularization of the survival curve in the Eastern European countries.* Paper presented at the Meeting of Young Demography, 2009, “Mortality and Morbidity Trends in Germany and Eastern Europe”. Rostock, Germany, 1.–2. 10. 2009. <http://www.young-demography.org/docs/pres_ST.pdf>.

THATCHER, A.R. 1999. The long-term pattern of adult mortality and the highest attained age. *Journal of the Royal Statistical Society. Series A* (Statistics in Society), 1999, Vol. 162, No. 1, pp. 5–43. <http://www.jstor.org/stable/2680465>.

THATCHER, A.R.; KANNISTO, V.; VAUPEL, J.W. 1998. *The Force of mortality at ages 80 to 120.* Odense University Press : Odense Monographs on Population Aging, 1998, Vol. 5. ISBN 87-7838-381-1.

VAN IMHOFF, E.; KEILMAN, N. 2000. On the quantum and tempo of fertility : Comment. *Population and Development Review*. September 2000, Vol. 26, No. 3, pp. 549–553.

VAUPEL, J.W. 1997. Trajectories of mortality at advanced ages. In: WACHTER, K.W.; FINCH, C.E. (eds.). *Between Zeus and the Salmon : Biodemography of Longevity*. Washington, D.C. : National Academy Press, 1997, pp. 17–37. ISBN 0-309-05787-6.

VAUPEL, J.W. 2002. Life expectancy at current rates vs. current conditions : A reflexion stimulated by Bongaarts and Feeney’s “How long do we live?”. *Demographic Research*. Vol. 7, No. 8, pp. 365–378. <http://www.demographic-research.org/Volumes/Vol7/8/7-8.pdf>.

VAUPEL, J.W. 2009. Lively questions for demographers about death at older ages. *Population and Development Review*. 2009, Vol. 35, No. 2, pp. 347–356.

VAUPEL, J.; MANTON, K.; STALLARD, E. 1979. The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography*. 1979, Vol. 16, pp. 439–454.

VAUPEL, J.W.; YASHIN, A.I. 1983. The deviant dynamics of death in heterogeneous populations. *Research Report.* 1983, Vol. 1, pp. 179–211. < http://user.demogr.mpg.de/jwv/pdf/Vaupel-SociologicalMethodology-15-1985.pdf >.

VAUPEL, J.W.; YASHIN, A.I. 1985. Heterogeneity’s ruses : Some surprising effects of selection on population dynamics. *The American Statistican.*1985, Vol. 39, No. 3, pp. 176–185.

VAUPEL, J.W.; YASHIN, A.I. 1986. Targeting lifesaving: Demographic linkages between population structure and life expectancy. *European Journal of Population.* 1986, Vol. 2, No. 3–4, pp. 335–360.

VAUPEL, J.W.; YASHIN, A.I. 1987. Repeated resuscitation : How lifesaving alters life tables. *Demography.* 1987, Vol. 24, No. 1, pp. 123–135. <http://user.demogr.mpg.de/jwv/pdf/Vaupel-Demography-24-1987-1.pdf >.

WACHTER, K.W. 1997. Between Zeus and the Salmon: Introduction. In: WACHTER, K.W.; FINCH, C.E. (eds.). *Between Zeus and the Salmon : Biodemography of Longevity*. Washington, D.C. : National Academy Press, 1997, pp. 1–16. ISBN 0-309-05787-6.

WACHTER, K.W. 2005. Tempo and its tribulations. *Demographic Research*. 2005, Vol. 13, No. 9, pp. 201–222.

WEDEN, M.M. 2007. Twentieth century U.S. racial inequalities in mortality : Changes in the average age of death and the variability in the age of death for white and non-white men and women, 1900–2002 . *Labor and Population, Working Paper.* 2007.

WEGNER, CH. 2010: Tempo effects in different calculation types of period death rates. *Comparative Population Studies*. 2010, Vol. 35, No. 3, pp. 543–568.

WIENKE, A. 2011. *Frailty Models in Survival Analysis*. Chapman & Hall/CRC biostatistics series, Taylor and Francis, 2011, 301 pp. ISBN 978-1-4200-7388-1.

WILMOTH, J.R. 1995. Are mortality rates falling et extremely high ages : An investigation based on a model proposed by Coale and Kisker. *Population Studies,* 1995, Vol. 49, No. 2, pp. 281–295.

WILMOTH, J.R. 1997. Search of limits. In: WACHTER, K.W.; FINCH, C.E. (eds.). *Between Zeus and the Salmon : The Biodemography of Longevity.* Washington, D.C. : National Academy Press, 1997, pp. 38–64. ISBN 0-309-05787-6.

WILMOTH, J.R.; HORIUCHI, S. 1999. Rectangularization revisited : Variability of Age at Death within Human Populations. *Demography.* 1999, Vol. 36, No. 4, pp. 475–495.

WILMOTH, J.R.; ANDREEV, K.; JDANOV, D.; GLEI, D.A. 2007. *Methods Protocol for the Human Mortality Database*. Human Mortality Database, May 31, 2007, Version 5. <http://www.mortality.org/Public/Docs/MethodsProtocol.pdf>.

YASHIN, A.I.; VAUPEL, J.W.; IACHINE, I.A. 1994. A duality in aging : The equivalence of mortality models based on radically different concepts. *Mechanisms of Ageing and Development*. 1994, No. 74, pp. 1–14.

YASHIN, A.I.; BEGUN, A.S.; BOIKO, S.I.; UKRAINTSEVA, S.V.; OEPPEN, J. 2001. The new trends in survival improvement require a revision of traditional gerontological concepts. *Experimental Gerontology.* 2001, No. 37, pp. 157–167.

YASHIN, A.I.; BEGUN, A.S.; BOIKO, S.I.; UKRAINTSEVA, S.V.; OEPPEN, J. 2002. New age patterns of survival improvement in Sweden : do they characterize changes in individual aging? *Mechanisms of Ageing and Development.* 2002, Vol. 123, No. 6, pp. 637–647. <http://www.sciencedirect.com>.

YI, Z.; LAND, K.C. 2001. A sensitivity analysis of the Bongaarts-Feeney new method for adjusting bias in observed period total fertility rates. *Demography*. 2001, Vol. 38, No. 1, pp. 17–28.

YI, Z.; LAND, K.C. 2002. Adjusting period tempo changes with an extension of Ryder's basic translation equation. *Demography*. May 2002, Vol. 39, No. 2, pp. 269–285.

YI, Z.; VAUPEL, J.W. 2003. Oldest-old mortality in China. *Demographic Research.* 2003, Vol. 8, Article 7, pp. 215–244. <http://www.demographic-research.org/volumes/vol8/7/8-7.pdf>.

ZHEN, Z.; VAUPEL, J.W. 2008. *The threshold between compression and expansion of mortality*. Paper presented at the Population Association of America 2008 Annual Meeting, New Orleans, April 17–19, 2008.

ZUREICK, S.M. 2010. *Certainty in timing of death: A new analysis of shifting mortality and life span disparity.* University of California, Berkeley, Dissertation Thesis, ProQuest Dissertations and Theses [cit. 2012-03-05], <http://search.proquest.com/docview/749725445>.