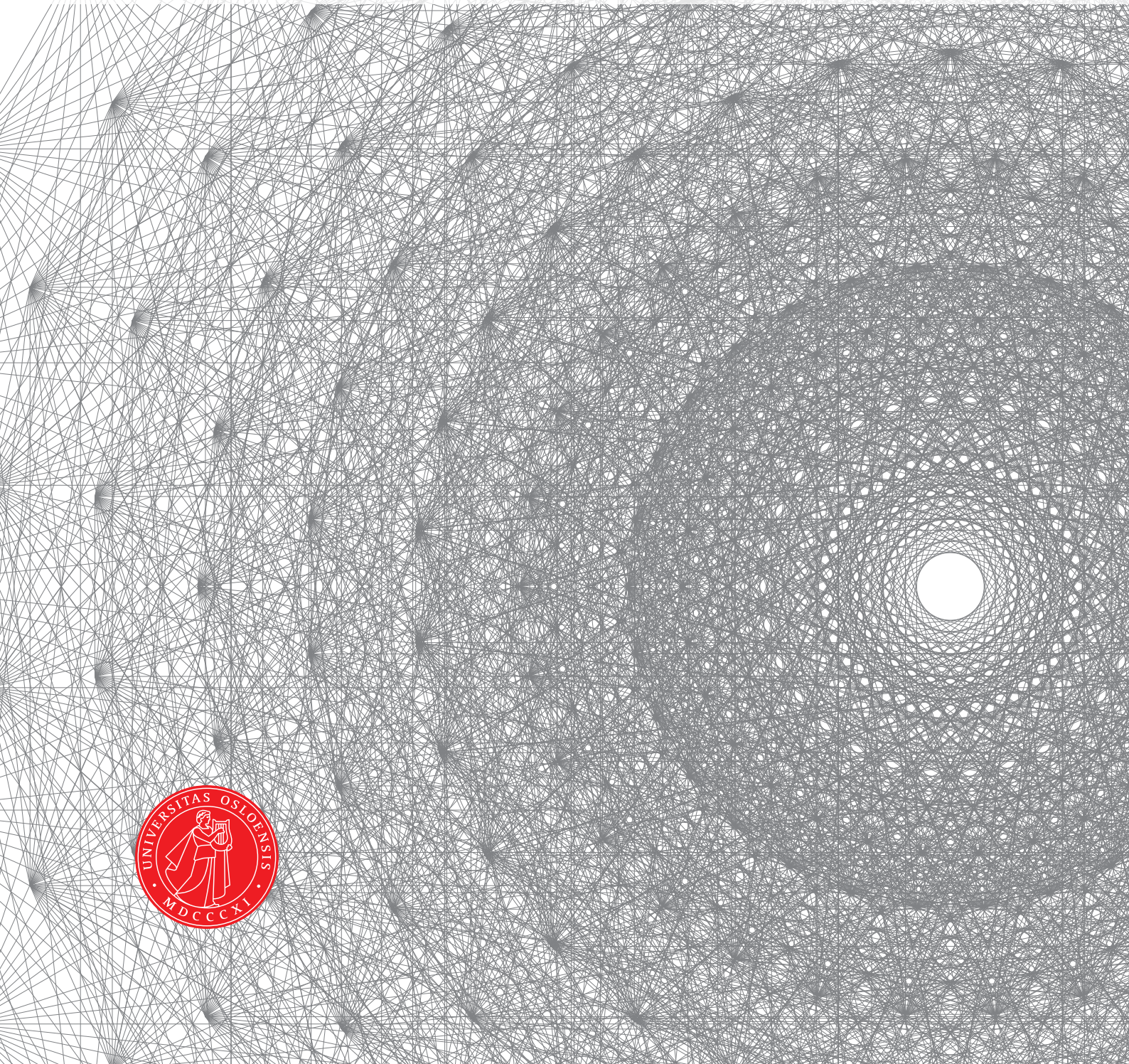
UiO : Department of Mathematics
University of Oslo

Comparison of Cohort and Period Life Expectancy Between Countries in Scandinavia and the Mediterranean

Joël Fomete Tankmo

Master's Thesis, Autumn 2019



This master's thesis is submitted under the master's programme *Stochastic Modelling, Statistics and Risk Analysis*, with programme option *Statistics*, at the Department of Mathematics, University of Oslo. The scope of the thesis is 60 credits.

The front page depicts a section of the root system of the exceptional Lie group E_8, projected into the plane. Lie groups were invented by the Norwegian mathematician Sophus Lie (1842–1899) to express symmetries in differential equations and today they play a central role in various parts of mathematics.

Abstract

Life expectancies at birth are usually computed from period life tables, i.e. on the basis of mortality in a particular calendar year. Over the last century, there has been a worldwide decline in mortality rates at most ages leading to an increase of the life expectancy. But comparing countries based on the life expectancy from period life tables may ignore different historical mortality developments in those countries and therefore lead to a wrong conclusion. Consequently, instead of comparing life expectancy for countries based on period life tables, it may be more appropriate do such comparison based on life expectancy from cohort life tables. Since cohort life expectancies can only be obtained for older cohorts i.e. those born more than a hundred years ago, in this thesis we suggest that for younger cohorts one may consider the expected number of years lost up to a given age. When we consider life expectancy based on period mortality, one finds that since the 80's, women in Spain and Italy have had higher life expectancy than those in Norway and Sweden. However, if we consider the expected number of years lost for different cohorts in Spain, Italy, Norway and Sweden, we observe that women in Scandinavia are still expected to lose fewer years, i.e. live longer, than those in the Mediterranean. The results indicated by period data may be due to a selection effect and may therefore be an artifact.

Acknowledgements

First I would like to express my sincere gratitude to my thesis supervisor Professor Ørnulf Borgan of the Faculty of Mathematics at the University of Oslo, for introducing me to an interesting topic and sharing his knowledge and experience during this past year. The door to Prof. Borgan's office was always open whenever I ran into a trouble spot or had a question about my research or writing.

The years at the University of Oslo would have been significantly less interesting without my fellow students. Thank you to Amine, Amanda, Maria and Siren. I appreciate all the insightful discussions and the memorable times we had together. I would also like to thank Prof. Frank Proske for the encouragement and support. I wish to express my thanks to Kari Mørk for the encouragement and the flexibility that was granted to me at Finance Norway during this time.

I must express my very profound gratitude to my family, especially to the Mehta's and to Harriet and Jacob Natvig for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. I would like to offer special thanks to Birma for her unceasing support and love.

I dedicate this thesis in memory of my dear parents Esther and Jonas Fomete, you are gone but made this journey possible. Lastly, and most of all, I would like to thank GOD, " For from him and through him and to him are all things. To him be glory forever. Amen." Rom 11:36

Blindern, October 24, 2019

Joël Fomete Tankmo

Contents

CHAPTER 1

INTRODUCTION

According to data from Statistics Norway, life expectancy at birth for Norwegian and Swedish women in 2017 was respectively 84.3 years and 84.1 years, while in Italy and Spain it was respectively 85.2 years and 86.1 years. But about 60 years earlier (in 1960), women in Scandinavia had higher life expectancy than women in the Mediterranean (Norway 76.0 years, Sweden 74.9 years, Italy 72.3 years and Spain 72.2 years). Thus, there has been a major change in life expectancy at birth over the last fifty years. (Sønstebø 2019)

Life expectancies at birth are usually computed from period life tables, ie. on the basis of mortality in a particular calendar year. When there are major changes in mortality, and these changes occur at different times in different countries, life expectancy calculated based on period life tables can give a misleading picture of life expectancy in different countries. It would then be better to calculate life expectancy based on a cohort, ie. from the mortality rates of persons who were born in the same year. However, a problem with the latter approach is that one cannot find the life expectancy of a cohort until the entire cohort has died out. A partial solution is to compute the expected number of years lost for the cohorts up to given ages. (Borgan and Keilman 2019)

As a cohort of people ages, the individuals at highest risk tend to die first. This differential selection can produce patterns of mortality for the population as a whole that are surprisingly different from the patterns for sub-populations or individuals. This can be illustrated by the use of frailty models. (Aalen, Borgan, and Gjessing 2008)

The data we will use are from the Human Mortality Database (HMD) which contains a wealth of information about detailed mortality and population data for about 40 countries. The input data consist of death counts from vital statistics, plus census counts, birth counts, and population estimates from various sources. Detailed information about mortality rates for different countries can reveal information about changes in life expectancy.

This thesis is organized into five main chapters. We will start in the second chapter by considering some basic concepts and results in survival analysis and also consider parametric inference for survival data. The third chapter contains information about Lexis diagram and how it is constructed. And then we will see how life line, age, period and cohort data are visualize in a Lexis diagram. We will define mortality rate and show how the exposure and the number of deaths are computed from the Lexis triangles. These are then used to compute the mortality hazard for period data and cohort data for Norwegian females from the calendar year 1900 to the calendar year 2014. We will plot the period mortality rates for Norwegian females during five different periods and similarly for the cohort mortality rates. We define and compute the life expectancy. We then plot the life expectancy at birth for period data. We also plot the life expectancy up to age a first as a function of cohort and then as a function age for different choices of a for Norwegian females. We end the second section by computing and plotting the expected number of years lost up to a certain age a for different cohorts of Norwegian females.

In the fourth chapter we compare the mortality for two countries in the Mediterranean and two countries in Scandinavia. We begin by computing and plotting the mortality rates for Italy, Spain, Norway and Sweden.To have uniform data sets for the four countries we choose to look at developments from the calendar year 1910 to the calendar year 2014.The ages varies between 0 and 110 years. We then compare the evolution of mortality in those countries over the years. The mortality rates are again use to compute and plot the period life expectancy for the years 1920, 1950, 1980 and 2014. We also compute the expected number of years lost where we first fix the age and consider the expected number of years lost as a function of cohort and then we fix the cohort and look at the expected number of years lost as a function of age.

The fifth chapter of the thesis focuses mainly on the frailty model. In the sixth chapter, we use the

two-points frailty distribution with a baseline death intensity of the Gompertz-Makeham form to illustrate the difference between Scandinavia and the Mediterranean and discuss a possible explanation of why period data and cohort data may lead to different results.

CHAPTER 2

SURVIVAL ANALYSIS BACKGROUND

The term survival data is commonly used to denote data that measure the time to some event. In this thesis, the event is death. The term is also used with other events, like the time to failure of a component in an unit, the occurrence of a disease or a complication. Usually to have an equal footing among the individuals, the time origin must be specified. e.g. the date of birth or the time point of diagnosis of a type of cancer. Survival data are generally described and modelled in terms of two related quantities, namely survival and hazard. (Clark et al. 2003)

In the following we will consider some basic concepts and results in the survival analysis set-up and also consider parametric inference for survival data taken from lecture notes in life history analysis by Borgan (1990).

2.1 Basic concept and results

We consider a survival time, i.e. positive random variable T with cumulative distribution function (c.d.f) $F(t) = P(T \leq t)$, and the probability density function (p.d.f) $f(t) = F'(t)$. We usually consider the survival distribution

$$S(t) = 1 - F(t) = P(T > t) \tag{2.1}$$

instead of the distribution function itself. Furthermore, we introduce the death intensity, or the hazard, defined as

$$\mu(t) = f(t)/S(t) \tag{2.2}$$

To get an alternative, and more directly interpretable expression for $\mu(t)$, note that

$$P(t < T \leq t + \Delta t | T > t) = [S(t) - S(t + \Delta t)]/S(t).$$

Hence using (2.1) and (2.2), we get

$$\mu(t) = -S'(t)/S(t) = \lim_{\Delta t \to 0} P(t < T \leq t + \Delta t | T > t)/\Delta t \tag{2.3}$$

This shows that for small values of Δt, $\mu(t)\Delta t$ equals approximately the probability of dying in $(t + \Delta t]$ for an individual who is still alive at age t. In this respect $\mu(\cdot)$ measures the instantaneous risk of dying. To express the survival distribution and the density in terms of the death intensity, we see that (2.3) yields

$$\mu(t) = -\frac{d}{dt} \ln S(t)$$

The well known formula

$$S(t) = e^{-\int_0^t \mu(v)dv} \tag{2.4}$$

results. Differentiating (2.4) we get

$$f(t) = \mu(t) e^{-\int_0^t \mu(v)dv} \tag{2.5}$$

The expected life length (life expectancy at birth) is given by

$$E[T] = \int_0^\omega t f(t) dt, \tag{2.6}$$

where ω is the highest possible age. Integrating by parts, and making use of the fact that $-f(t)$ is the derivative of $S(t)$, which has limits or boundary conditions $S(0) = 1$ and $S(\omega) = 0$, one can show that:

$$E[T] = \int_o^\omega S(t)dt \tag{2.7}$$

In words, the expected life length is simply the integral of the survival function. We will also be interested in the expected number of years lived up to a given age a. We introduce therefore :

$$T_a = \min(T, a)$$

and find that the expected number of years lived up to age a is :

$$E[T_a] = \int_0^a tf(t)dt + aP(T > a) = \int_0^a S(t)dt$$

Finally we will also look at the expected number of years lost up to age a . It is given as :

$$a - E[T_a] \tag{2.8}$$

2.2 Piece-wise constant mortality

In this section we will look specially at the situation where mortality is constant over one year intervals. We suppose that mortality can be given by:

$$\mu(t) = \sum_{j=0}^{\omega-1} \mu_j I_j(t) \tag{2.9}$$

where ω is the highest possible age and

$$I_j(t) = \begin{cases} 1 & \text{for } j \le t < j+1 \\ 0 & \text{otherwise;} \end{cases}$$

For $k \le t < k+1 \le \omega$ the survival function can be given as :

$$S(t) = \exp(-\int_0^t S(u)du) = \exp\,(\sum_{j=0}^{k-1} \mu_j - (t-k)\mu_k) \tag{2.10}$$

The expected life length becomes

$$\begin{aligned}
E[T] &= \int_0^\omega S(t)dt = \sum_{k=0}^{\omega-1} \int_k^{k+1} S(t)dt \\
&= \sum_{k=0}^{\omega-1} \int_k^{k+1} \exp\,(-\sum_{j=0}^{k-1} \mu_j - (t-k)\mu_k)dt \\
&= \sum_{k=0}^{\omega-1} \exp\,(-\sum_{j=0}^{k-1} \mu_j) \exp(k\mu_k) \int_k^{k+1} \exp\,(-t\mu_k)dt \\
&= \sum_{k=0}^{\omega-1} \exp\,(-\sum_{j=0}^{k-1} \mu_j) \exp\,(k\mu_k) \frac{1}{\mu_k}[-\exp\,((k+1)\mu_k) + \exp\,(-k\mu_k)] \\
&= \sum_{k=0}^{\omega-1} \exp\,(-\sum_{j=0}^{k-1} \mu_j) \frac{1}{\mu_k}(1 - \exp\,(-\mu_k))
\end{aligned}$$

When a is an integer we also find that

$$E[T_a] = \int_0^a S(t)dt = \sum_{k=0}^{a-1} \int_k^{k+1} S(t)dt$$
$$= \sum_{k=0}^{a-1} \exp\left(-\sum_{j=0}^{k-1} \mu_j\right) \frac{1}{\mu_k}(1 - \exp\left(-\mu_k\right))$$

2.3 Maximum likelihood estimation

Suppose we have a population with n individuals and that we observe the individual i from the age V_i to the age T_i. Here T_i corresponds to the time of death or censoring. Let $\delta_i = 1$ if T_i corresponds to the time of death and $\delta_i = 0$ if it corresponds to censoring. We suppose that the hazard $\mu(t)$ is dependent on a vector of parameters $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$ and write $\mu(t) = \mu(t; \boldsymbol{\theta})$. We want to derive the maximum likelihood estimator (ML-estimator) for $\boldsymbol{\theta}$. The likelihood corresponding to individual i is given as :

$$\Lambda_i(\boldsymbol{\theta}) = \mu(T_i; \boldsymbol{\theta})^{\delta_i} \exp\left(-\int_{V_i}^{T_i} \mu(u, \boldsymbol{\theta})du\right) \tag{2.11}$$

and the total likelihood is

$$\Lambda_i(\boldsymbol{\theta}) = \prod_{i=1}^{n} \mu(T_i; \boldsymbol{\theta})^{\delta_i} \exp\left(-\int_{V_i}^{T_i} \mu(u, \boldsymbol{\theta})du\right) \tag{2.12}$$

The ML-estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$ is the value of $\boldsymbol{\theta}$ which maximizes (2.11), or equivalently maximizes

$$\ln \Lambda(\boldsymbol{\theta}) = \sum_{i=1}^{n} \delta_i \ln \mu(T_i; \boldsymbol{\theta}) - \sum_{i=1}^{n} \int_{V_i}^{T_i} \mu(u; \boldsymbol{\theta})du \tag{2.13}$$

Now $\hat{\boldsymbol{\theta}}$ is found by solving the set of equations

$$\frac{\partial \ln \Lambda(\boldsymbol{\theta})}{\partial \theta_j} = 0 \qquad j = 1, ..., p;$$

and checking that the solution of these equations yields a maximum of (2.13).

We then suppose that the mortality hazard is piece-wise constant over one year intervals as in (2.9), here we have $\boldsymbol{\theta} = (\mu_0, \mu_1, \ldots, \mu_{\omega-1})$. The part of the likelihood corresponding to individual i may now be written as

$$\Lambda_i = \Lambda_i(\boldsymbol{\theta}) = \mu(T_i; \boldsymbol{\theta})^{\delta_i} \exp(-\int_{V_i}^{T_i} \mu(u; \boldsymbol{\theta})du)$$
$$= \mu_1^{D_{i0}} \mu_2^{D_{i1}} \mu_2^{D_{i2}} ... \mu_{w-1}^{D_{i(w-1)}} \exp(-\sum_{k=0}^{w-1} T_{ik}\mu_k),$$

where

$$D_{ik} = \begin{cases} 1 & \text{for } k \leq T_i < k+1, \quad \delta_i = 1; \\ 0 & \text{otherwise;} \end{cases}$$

$$T_{ik} = \begin{cases} 0 & \text{if } V_i > k+1 \text{ or } T_i < k, \\ k+1-V_i & \text{if } k \leq V_i < k+1 \leq T_i, \\ T_i - V_i & \text{if } k \leq V_i < T_i \leq k+1, \\ T_i - k & \text{if } V_i < k \leq T_i < k+1, \\ 1 & \text{if } V_i < k \text{ and } T_i \geq k+i; \end{cases}$$

We see that D_{ik} is the number of times we observe a death for the ith individual in $[k, k+1)$, while T_{ik} is the total time the ith individual is observed to live in this interval.

Now the total likelihood is:

$$\Lambda = \prod_{i=1}^{n} \Lambda_i = \prod_{i=1}^{n} [\prod_{k=0}^{\omega-1} (\mu_k^{D_{ik}}) \exp(-\sum_{k=0}^{\omega-1} T_{ik}\mu_k)] = \prod_{k=0}^{\omega-1} (\mu_k^{D_k}) \exp(-\sum_{k=0}^{\omega-1} R_k\mu_k),$$

where $D_k = \sum_{i=1}^{n} D_{ik}$ is the observed number of deaths in $[k, k+1)$, and $R_k = \sum_{i=1}^{n} T_{ik}$ is the total exposure in this interval.

Thus we have

$$\ln \Lambda = \sum_{k=0}^{\omega-1} D_k \ln \mu_k - \sum_{k=0}^{\omega-1} R_k\mu_k,$$

and we find that

$$\frac{\partial \ln \Lambda}{\partial \mu_k} = \frac{D_k}{\mu_k} - R_k.$$

It follows that the ML-estimator for μ_k is

$$\hat{\mu}_k = \frac{D_k}{R_k} \tag{2.14}$$

i.e an occurrence/exposure rate.

These concepts and results in the survival analysis and parametric inference for survival data that are presented in this chapter will be used through out the rest of the thesis. In the next chapter, we will introduce the Lexis diagram and show how the mortality rates are computed from the Lexis diagram.

CHAPTER 3

COHORT AND PERIOD DATA

3.1 Data description

Our analysis are based on data from the Human Mortality Database (HMD, Last modified: 26 Sep 2017), which are freely accessible at http:// www.mortality.org. The HMD contains aggregate mortality statistics such as death counts, population estimates, exposure to risk estimates, life tables as well as some other statistics. Input data files for more than 35 countries, are accessible from each country page. In the HMD, every country is identified by a specific code. For example "NOR" identifies the national population of Norway . We used " Deaths by Lexis triangles " and " Exposure-to-risk by Lexis triangles " as describe later in connection with figures 3.8 and 3.10.

At the time of writing, the following periods, age and cohorts were available for Norway: For the period data we have information for all periods, from the calendar year 1846 to the calendar year 2014 and the age varying between 0 and 110+. For the cohort data we have information for all ages until the age of the cohort in 2014.(e.g for the cohort born in 1960 we have information up to the age of 54.)

3.2 Construction of the Lexis diagram

Any dynamics such as births and deaths involve change over calendar time, age, and/or cohort. Those dynamics can be visualized with the help of the Lexis diagram, named after the German statistician, economist, and social scientist Wilhelm Lexis (1837-1914). For a review of Lexis diagram, see e.g. Carstensen (2007).

The lexis diagram consist of a Cartesian coordinate system where the calendar time ("period") is depicted on the x-axis and age on the y-axis. Every demographic event can therefore be located based on the time and age. However, the Lexis diagram can not be resume to a Cartesian diagram because of two characteristics (Vandeschrick 2001):

- The Lexis diagram has two axes but allows the use of three separate coordinates.
- Moreover, on the Lexis diagram each individual under observation belong to a forced trajectory from which he can not escape, namely his 'Life line'.

The life line is a specificity of Lexis diagram and it is essential for defining cohort. Generally events that are observed on a Lexis diagram can be classified according to three coordinates. For example death will have:

- The date of death (time designated by T),
- The age of the person at the time of death (age designated by A),
- and the moment of birth of the person (designated by M).

On the Figure 3.1 we have a lexis grid from year 1990 to year 1995, representing the age 0 to 5. The units on those axes are identical. The moment of birth for each individual correspond to the age of exactly 0 year. The diagram is divided into 1 x 1 cells (i.e one year of age by one year of time). A simple mathematical relationship connects the three variables time, age and time of birth : A = T - M. Since these three variables are expressed in the same unit (year), all the life lines will have the same slope, drawing a 45 degrees angle with the time of birth axis.

Figure 3.2 is a Lexis diagram with the straight red lines representing "life line". As we can observe, the line begins on the time axis at the person's birth. The line is continuous and the ends point represent

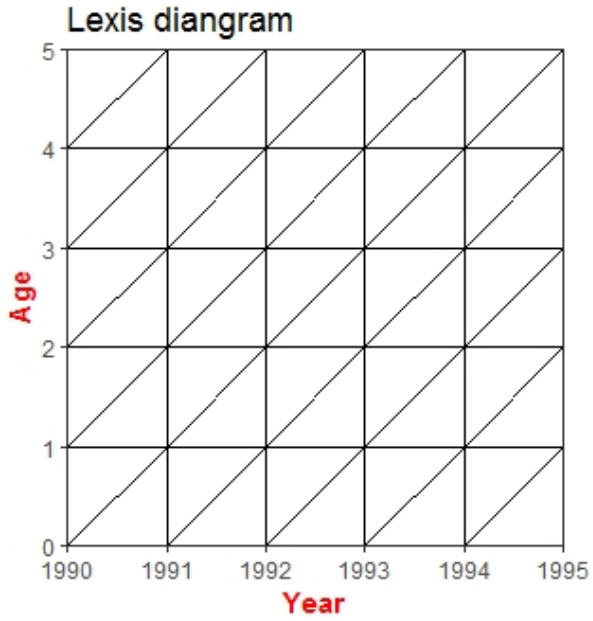
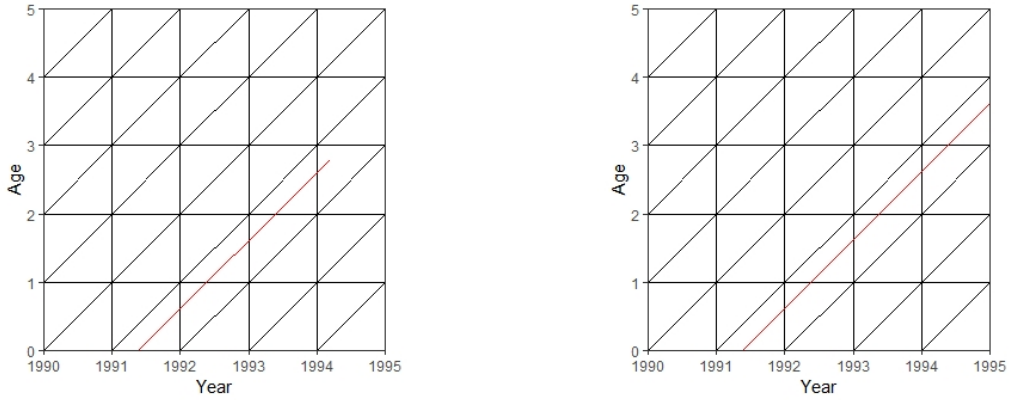


Figure 3.1: Lexis grid



(a) (b)

Figure 3.2: Lexis diagram with life line

the person's death. So in the figure on the left we have a person born the 23 of September 1991 and died on the 11 of June 1994. On the right we have the representation of a person born the 23th of September 1991 and still alive at the end of 1994. If we add together all the life line lengths in a particular portion of the Lexis diagram, we will obtain the person-years lived or exposure in that area.

A cohort can be defined as a group of persons who experience an event at the same time. On Figure 3.3 we have a Lexis diagram presenting a cohort of persons born in the year 1990 and lived until the end of 1994. The Lexis diagram can present a cohort through their life experience or a cohort in a specific interval.

Many demographic investigations are conducted on period data. On Figure 3.4 we have the age group from 0 year to 5 years during the year 1992, i.e population during year 1992. Finally, on Figure 3.4 we have highlighted all points that belong to the age of 2 years.

In summary, demographers use the term "diagram" to refer to their graphical representation of the data rather than "graph". On the Lexis diagram, the x-axis and the y-axis usually support the calendar time and the age respectively. On this diagram it is possible to identify events according to two coordinates on the axes such as it is the case in any type of Cartesian diagram, but the Lexis diagram has the specificity of having a third coordinate, the moment of birth. (Wilmoth et al. 2007)

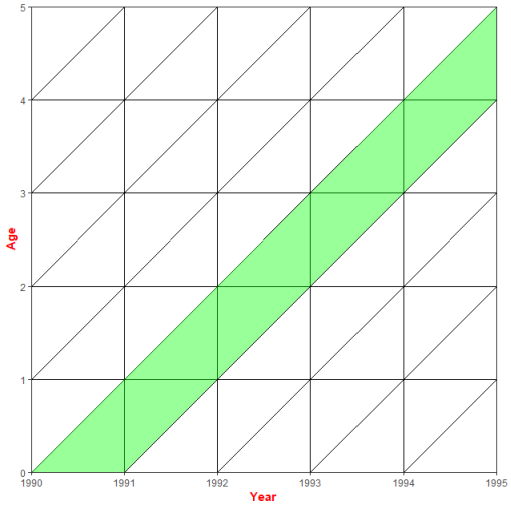


Figure 3.3: Cohort from 1990

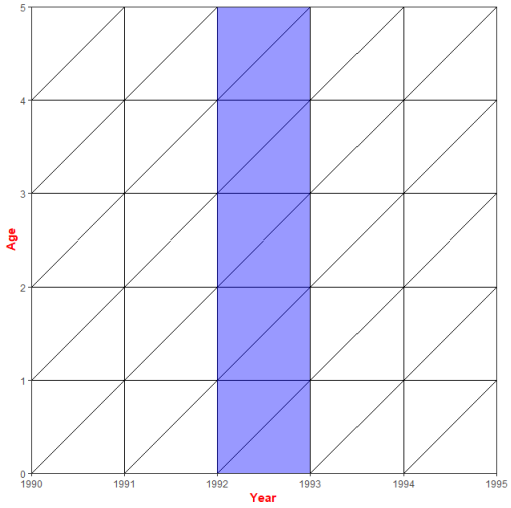


Figure 3.4: Lexis diagram with year 1992

3.3 Mortality rate

One way of understanding population change is to measure and analyse its components. Demographers therefore measure events in terms of rates. A mortality rate $\hat{\mu}$ can be defined as the ratio of the number of deaths (D) in a specified time period by the exposure (i.e person-years) during the period (R):

$$\hat{\mu} = \frac{D}{R} \tag{3.1}$$

As shown previously in (2.14), this is the maximum likelihood estimate of the mortality hazard. We will first look at how the mortality rate is computed in the case of period data and then the cohort data.

3.3.1 Period data

An individual age x, where x is an integer, has an exact age within the interval $[x, x+1)$. This concept is called "last birthday". Similarly an event that occurs in year t occurs during the calendar time interval $[t, t+1)$. The same apply to the exposure to risk of dying. The person-years age x in year t refers to all person-years lived in the age interval $[x, x+1)$ during calendar time $[t, t+1)$. We assume that mortality hazard is constant over each one-year age interval and calendar year (period) and denote its value for age x and year t by $\mu_{x,t}$ see figure 3.7. The mortality hazard may be estimated by:

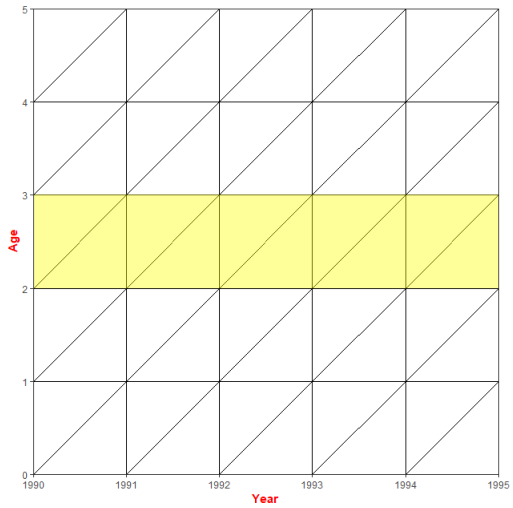


Figure 3.5: Lexis diagram with age group 2

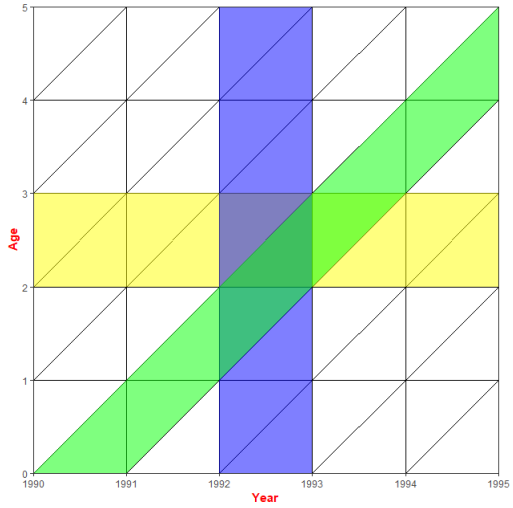


Figure 3.6: Summary of Lexis with age, year and cohort

$$\hat{\mu}_{x,t} = \frac{D_{x,t}}{R_{x,t}} \tag{3.2}$$

where $D_{x,t}$ is the number of deaths at age x in year t and $R_{x,t}$ the exposure (i.e person-years) at age x in year t.

If the coordinates of all life-lines are known, then the exposure $R_{x,t}$ can be calculated precisely by adding up the length of each line segment within the cell. The actual length of each segment must be divided by $\sqrt{2}$, since life-lines are 45° from the age or time axes. However in studies of large national populations we do not need to make calculations to obtain $D_{x,t}$ and $R_{x,t}$ because they may be obtained from the Human Mortality Database (HMD). More specifically, what may be obtained from the HMD are the number of deaths and the exposures (i.e person years) for the triangles in lexis diagram. We denote by $D^{(l)}_{x,t}$ and $D^{(u)}_{x,t}$ the number of deaths in the lower and upper triangles for age x in year t, see figure 3.8 and let $R^{(l)}_{x,t}$ and $R^{(u)}_{x,t}$ be the corresponding exposures.

In order to compute the mortality rate (3.1) for the period data, we then add up the upper (blue) and the lower (yellow) Lexis triangle of the same cell. See figure 3.8. This gives

$$D_{x,t} = D^{(u)}_{x,t} + D^{(l)}_{x,t} \tag{3.3}$$

and similar for the exposure:

$$R_{x,t} = R^{(u)}_{x,t} + R^{(l)}_{x,t} \tag{3.4}$$

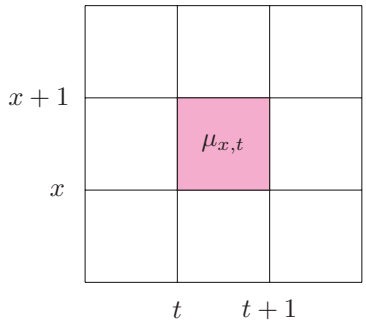


Figure 3.7: Mortality hazard for age x and year t.

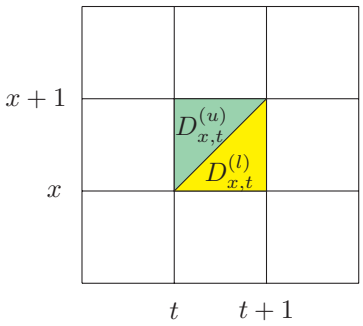


Figure 3.8: lower and upper triangle.

3.3.2 Cohort data

For cohort data we assume constant mortality hazard in the parallelogram in figure 3.9 and denote its value for cohort c in age x by $\mu_{x,c}$. In this case, we add together the deaths in the yellow and green triangles of figure 3.10 to obtain the number of deaths at age x for the cohort born in year c. The cohort death rates is then estimated by:

$$\hat{\mu}_{x,c} = \frac{D_{x,c}}{R_{x,c}} \tag{3.5}$$

where

$$D_{x,c} = D^{(l)}_{x,c+x} + D^{(u)}_{x,c+x+1} \tag{3.6}$$

and similar for the exposure:

$$R_{x,c} = T^{(l)}_{x,c+x} + R^{(u)}_{x,c+x+1} \tag{3.7}$$

3.4 Plots of mortality rate

The graph in figure 3.11 presents the mortality rates for Norwegian females at each age for period data during the years 1900, 1930, 1960, 1990 and 2014. The mortality rates are presented here on logarithmic scale "per 1000" person years. We can observe that the mortality rates are quite high just after birth and the first year of life. But the overall trend is that after this the rates of dying fall gradually, attaining minimum risk at age 5 for the year 2014 (graph in blue) and 15 for the year 1900 (graph in red). Then the risk starts increasing in adolescence, we can observed an exponential rise from one age year to the next. From the age 100 and above, it may be difficult to have an accurate estimation because of the very low number of persons alive and therefore low number of deaths.

The graph in figure 3.12 presents the mortality rates for females at each age for cohort data for the cohorts born in 1900, 1920, 1940, 1960 and 1980 in Norway. As for the period data, we can observe here that the risk of dying is quite high immediately after birth but falls considerably reaching a minimum risk around 10 year for the cohort 1980 (graph in blue) and 15 for the cohort 1900 (graph in red). Between the age 18 and 43 we can observe some small fluctuations without any significant increase or decrease of the risk of dying. Around age 50 we can observe an exponential rise from one year to the next of the risk of dying. We also observe that the cohorts 1960 (graph in gold) and 1980 (graph in blue) stop at ages 54 and 34 respectively. That is due to the fact that we don't have data after that age for these cohorts. As seen in graphs in figures 3.11 and 3.12 we can conclude that there is a continuous decrease of the risk of dying from year to year for all age groups. This mortality decline could be a result of an improvement of healthy lifestyle and diet during the past years and the development of new drugs and techniques in the medical field.

3.5 Life expectancy

Life expectancy in any given year can be defined as the average number of years a person born in that year is expected to live if mortality rates at each age were to remain the same in the future. The life expectancy can be shown separately for males and females, as well as a combined figure. In figures 3.13 and 3.14 we focus on Norwegian females. Life expectancy can be used as a measure or indicator of the

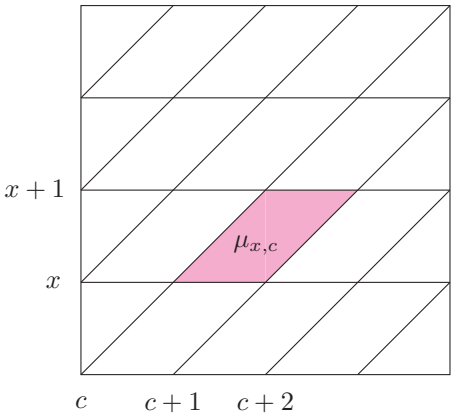


Figure 3.9: Mortality hazard for age x and Cohort c.

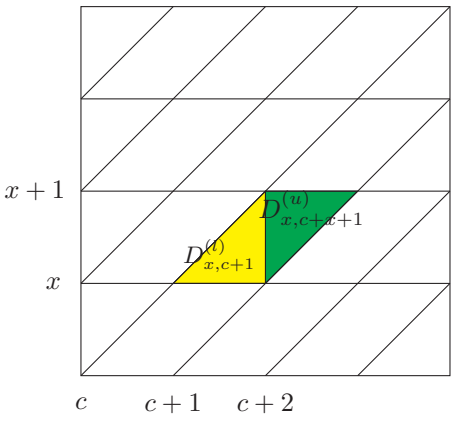


Figure 3.10: lower and upper triangle

quality of healthcare in a country, an ongoing war, or a pandemic. Figure 3.13 shows the evolution of the life expectancy at birth for Norwegian females in the period 1900-2014. In 1900, life expectancy was 55.09 years and 84.09 years for 2014. We observe that life expectancy has increased more or less continuously over the years, but in 1918 we can see a clear interruption due to the Spanish flu pandemic. It went from 59.05 years two years earlier (1916) to 52.03 years in 1918, that is a loss of about 7 years in life expectancy in a very short period of time. We also observe a little fluctuation from year to year during the period 1900-1916 and a small decrease during the world wars.

Figure 3.14 shows the life expectancy up to a certain age a for different choices of a for cohorts of Norwegian females. In this case we fix the age a and consider life expectancy as a function of the cohort c. The cohorts vary between 1900 and 2010. The life expectancy up to 100 years for the cohort born in 1900 was 62.43 years and 69.93 years for the cohort born in 1915. That is an increase of 7.05 years. The evolution for that age group can be observed on the plot in red. The life expectancy up to 80 years for the cohort born in 1920 was 68.35 years and 71.97 years for the cohort born in 1935. Between 1920 and 1935 we observe an increase of 3.62 years (magenta line). The increase of life expectancy for age 100 and 80 and their various cohorts is due to the drop in mortality we saw in figure 3.12. For the age 60 and cohorts born in 1940, 1950 and 1955, we obtain respectively 55.95 years, 57.54 years and 57.94 years. That can been observed on the green line. On the line in gold, with age 40 and cohorts 1960, 1970 and 1975, we obtain respectively 39,06 years, 39,36 years and 39,42 years. For the, age 20 and cohorts 1980, 1990 and 1995 we obtain respectively 19.82 years, 19.85 years and 19.90 years (blue line). The life expectancy up the ages 60, 40 and 20 and their various cohorts have a small increase. Most of the individual in those different cohorts are still alive while the individuals for the cohort 1900, 1920 and 1935 are all death or almost all death.

100 years is the maximum age, that means we will need 100 years of data for each cohort between 1900 and 2010. But that is not possible since very few cohorts have 100 years of data. To avoid this problem we have above used the "partial" life expectancy, that i.e we take the number of years lived up to a certain age a (e.g. 100, 80, 60, 40, 20 ...).

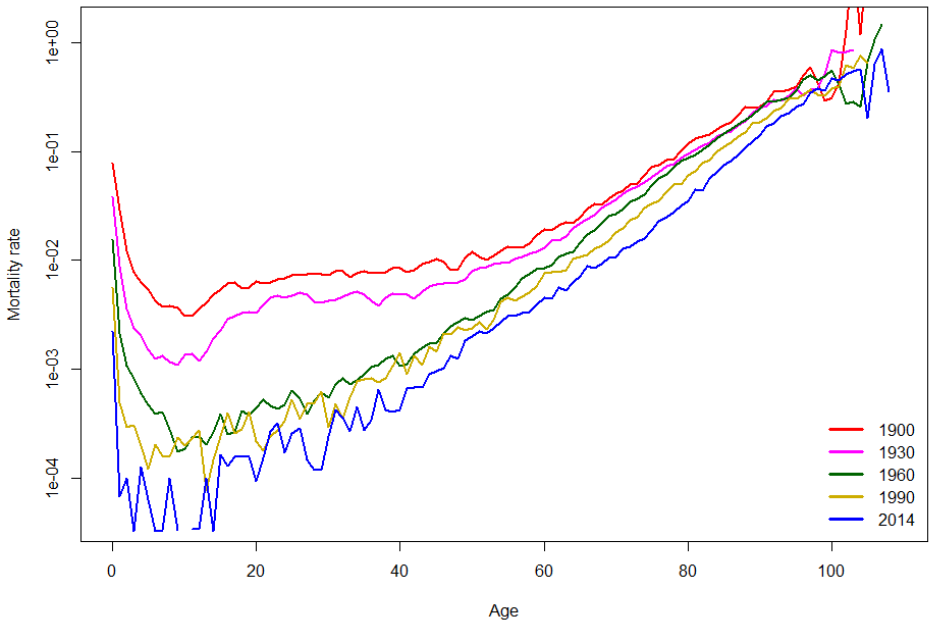


Figure 3.11: Period mortality rate for Norwegian females.

If there had been no mortality, the partial life expectancy up to age a would have been equal to a years. When there is mortality, the difference between age a and the partial life expectancy up to age a is called the expected number of years lost up to age a:

$$a - E[T_a]$$

On the figures 3.16 and 3.17 we have the plots of the expected number of years lost up to age a, first as a function of cohort for given age and then as a function of age for given cohorts for Norwegian women. We can observe on the figures that for all ages the number of years lost is lower for the more recent cohorts than the earlier ones. In the next chapter, we will apply the same methodology to some countries in Scandinavia and in the Mediterranean and compare the evolution of life expectancy in those two regions.

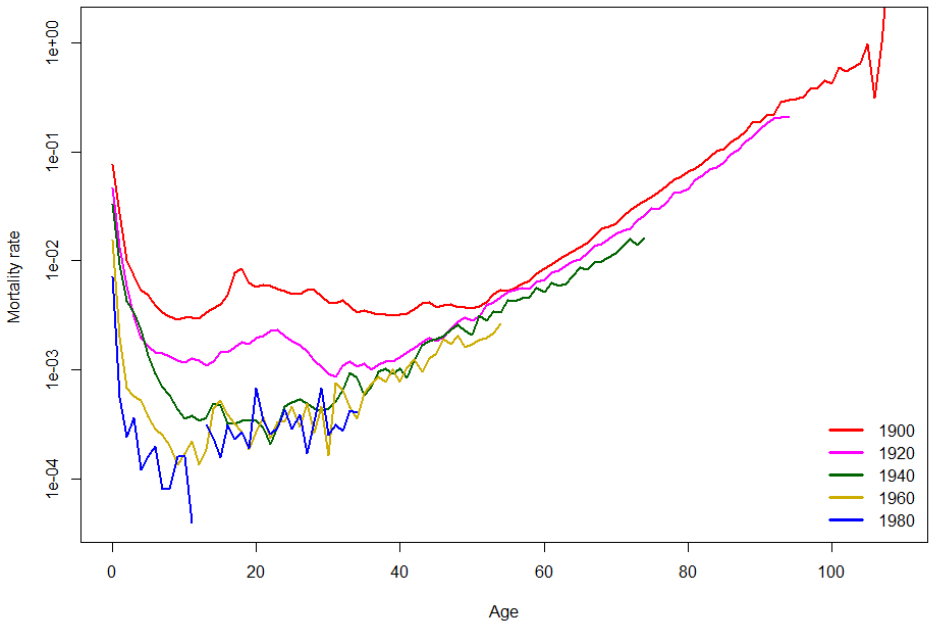


Figure 3.12: Cohort mortality rate for Norwegian females.

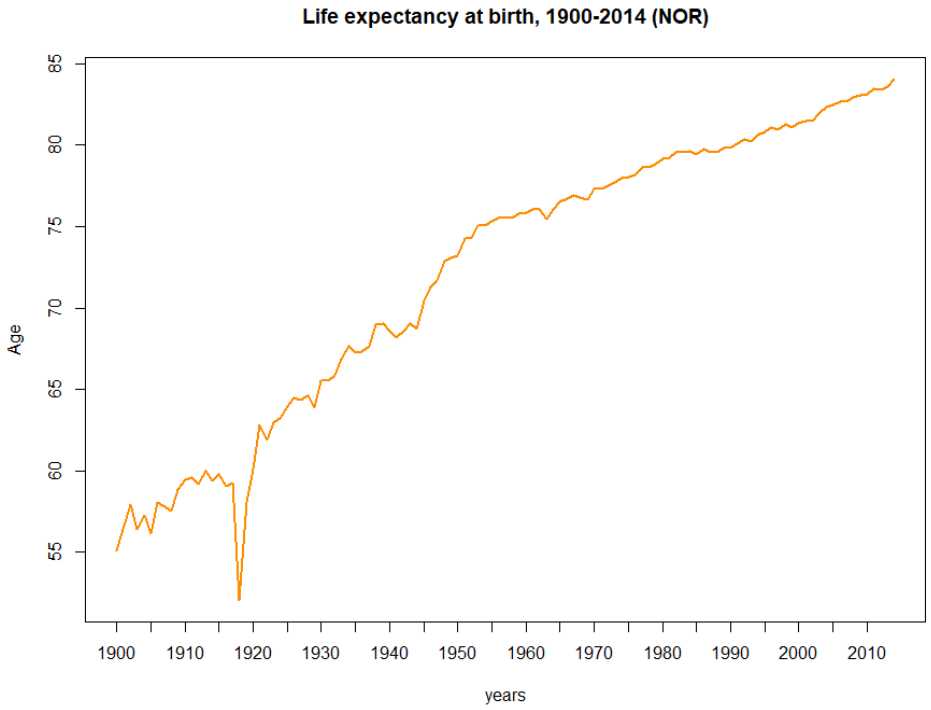


Figure 3.13: Life expectancy at birth for Norwegian females, period data.

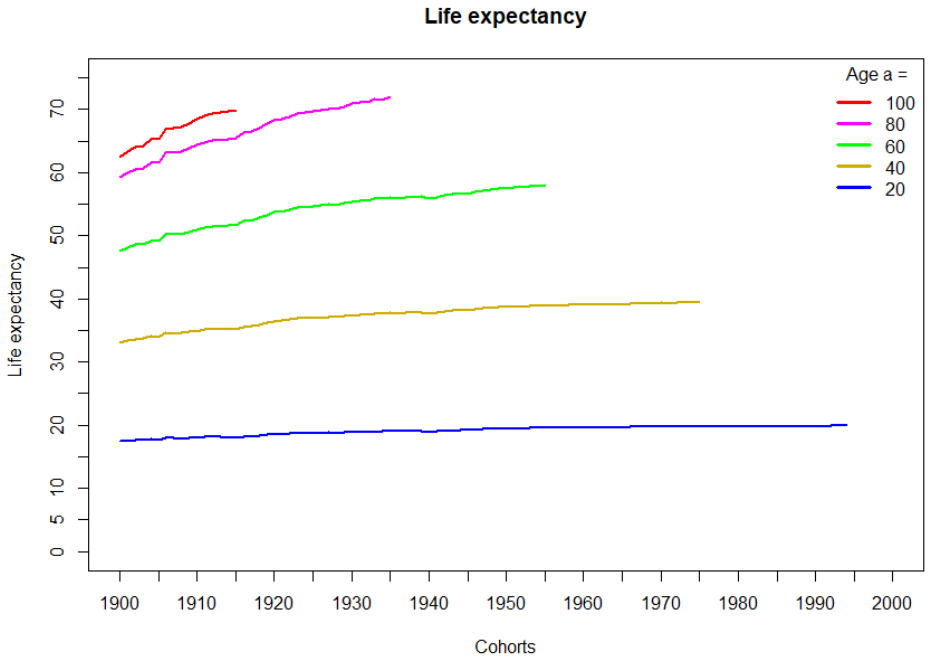


Figure 3.14: Life expectancy up to age a as a function of cohort for different choices of a for Norwegian females.

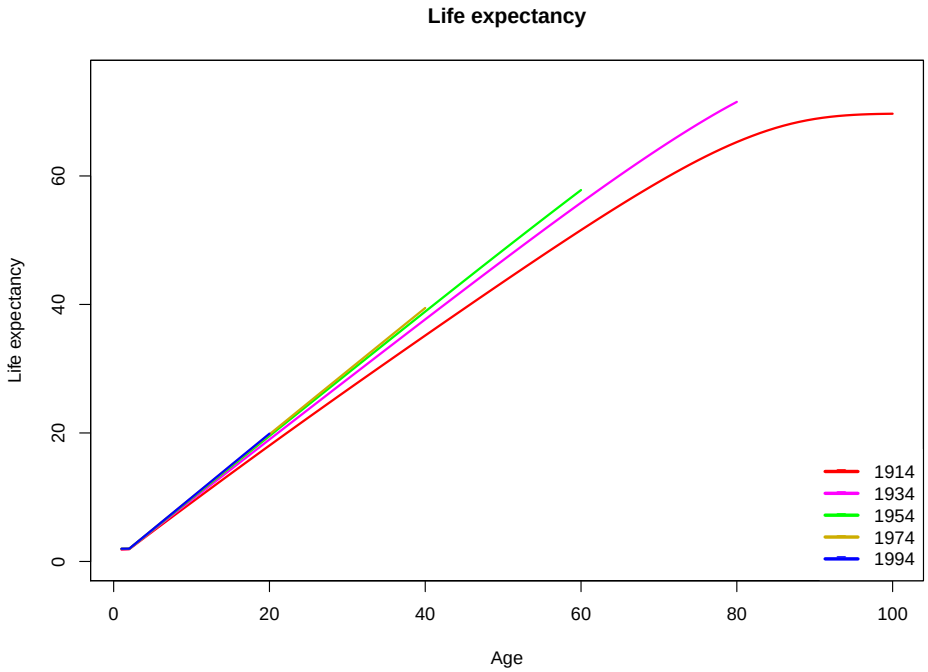


Figure 3.15: Life expectancy up to age a as a function of a for different cohorts of Norwegian females.

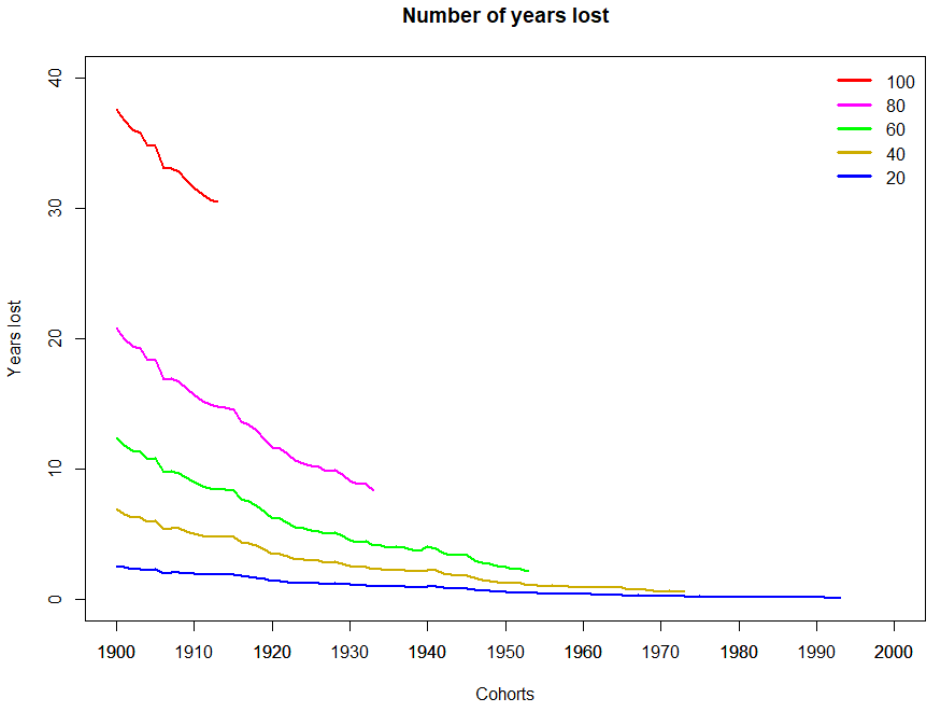


Figure 3.16: Expected number of years lost up to age a as a function of cohort for different choices of a for Norwegian females.

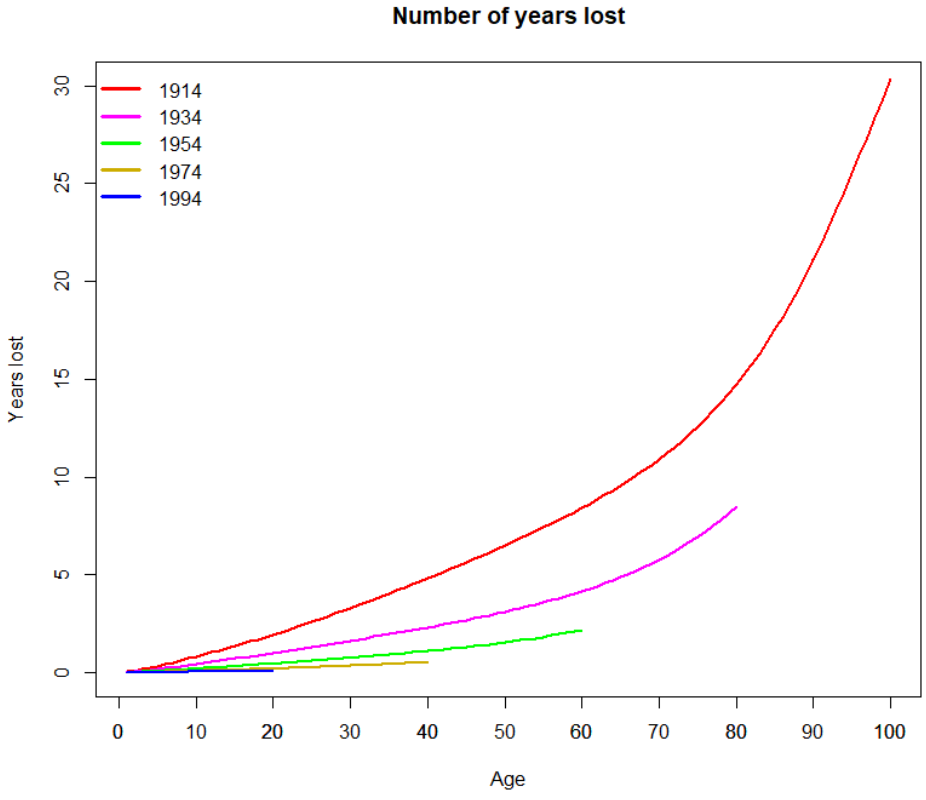


Figure 3.17: Expected number of years lost up to age a as a function of a for different cohorts of Norwegian females.

CHAPTER 4

COMPARING THE COUNTRIES

In this chapter, we want to compare four countries: Italy, Spain, Sweden and Norway. That is two countries from the Mediterranean and two countries from Scandinavia. The availability of the data in the Human Mortality database varies for these countries. The mortality series are as follow: in Norway we have data for 1846-2014, Sweden 1751-2017, Italy 1872-2014 and Spain 1908-2016. To have uniform data sets for the four countries we choose to look at developments from the calendar year 1910 (since mortality data for Spain are available only from the year 1908) to the calendar year 2014 (since Norway and Italy have mortality data until only 2014). The ages varies between 0 and 110 years.

4.1 Period data

4.1.1 Period Mortality

Over the last century, countries across the Mediterranean and Scandinavia have seen a lot of variation in mortality. In this section we will see how the period mortality has change during the years. The visualization from figure 4.1 to figure 4.4 shows the evolution of the period mortality rates in Italy, Spain, Sweden and Norway during the years 1920, 1950, 1980 and 2014. For all four countries, we can observe that the mortality rate is generally very high from birth but decrease drastically and attain a minimum level around age 10. During the adolescence until around the age of 30, we can observe a stagnation of the mortality. From age 35 years and beyond, we observe an exponential increase of mortality. We also note that for all ages the mortality has decreased over time. (Note that the scale of the y-axis differs from figure 4.1 to figure 4.4).

Spain (plotted in green on the figures) has the highest mortality in 1920 and 1950, followed by Italy. Figures (4.1 and 4.2) Sweden and Norway have the lowest mortality and their plots almost overlap during the period 1920. We also observe a high variation of mortality in Spain during that period.

In the ages 0-55 years there is a gap between the four countries, but after that we observe that Italy and Spain catch up Sweden and Norway, having almost the same mortality. Approaching the age 80, we can see on figures 4.1 and 4.2 that Spain has the lowest mortality rate in 1920 and 1950.

Figure 4.3 shows that during the calendar year 1980 the mortality is very low for all the four countries, Norway (plotted in blue) has the lowest mortality and reaches the lowest level at age 10 on figure 4.3.

Figure 4.4 shows that there is no major difference of mortality between the four countries during the period 2014. We observe small jumps between ages 0-40 years for all four countries. This is due to the use of logarithmic scale on the y-axis and a very low mortality for the low ages. We observe here that the mortality in Spain and Italy is slightly lower than the mortality in Sweden and Norway around the age 45 and upward.

We can conclude that the trend over the years remains almost the same for Sweden and Norway on one side and for Spain and Italy on the other side. The mortality in the earlier period was higher in Spain and Italy than in Sweden and Norway, but from the 80s as figure 4.3 illustrates, the gap has gradually reduced as Spain and Italy have surpassed Sweden and Norway as we can observe in figure 4.4.

4.1.2 Period life expectancy

The life expectancy for the period data are based on the period mortality as we have shown in the figures 4.1 to figure 4.4 for the years 1920, 1950, 1980 and 2014.

Figure 4.5 shows the period life expectancy at birth for Italy, Spain, Sweden and Norway from the calendar year 1910 to the calendar year 2014. We observe a sudden decrease of life expectancy for all the four countries around the period year 1918. This sudden decrease could be the result of the influenza

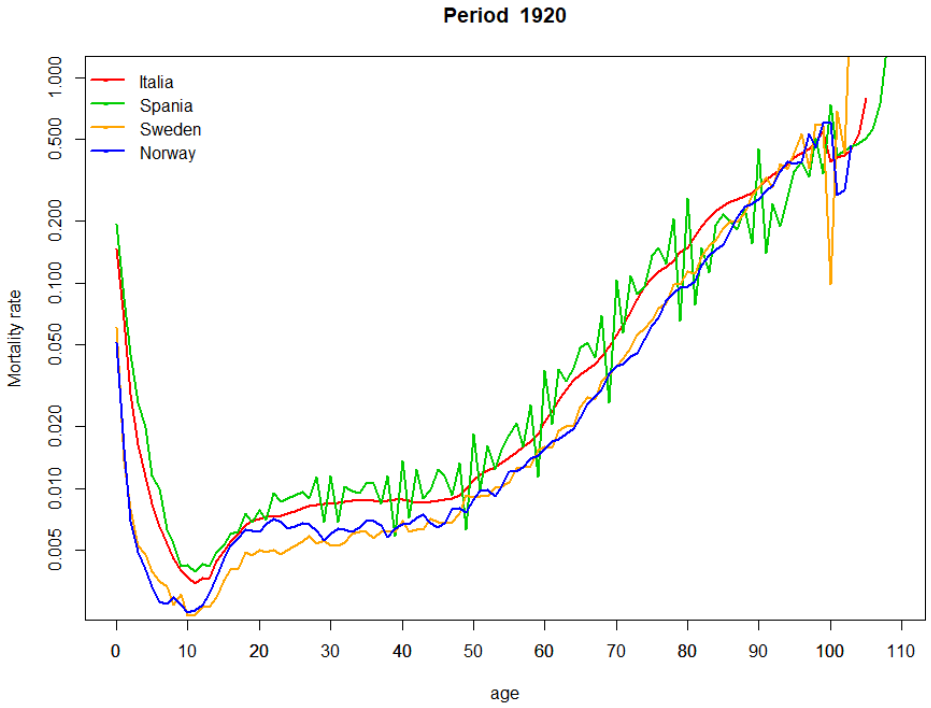


Figure 4.1: Period mortality rate for 1920 for all countries.

pandemic of 1918-1919 also known as the Spanish flu. The outbreak of this influenza virus, spread with astonishing speed around the world, killing millions of people. That may be why the effect is visible in all the four countries. The plots show that during that period Italy and Spain had the lowest life expectancy around 26 and 28 years respectively, while during the same period Sweden and Norway had about the same life expectancy around 52 years. Between the two groups we can notice a very big difference of life expectancy, about 24 years. We also observe a little drop in life expectancy around the calendar years 1938 and 1943 in Spain and Italy respectively, but this time Sweden and Norway remain stable. This is probably the result of the Spanish civil war and the world war II. Apart from the irregular pattern observed in 1918 for all the four countries and during 1938-1943 for Spain and Italy, we observe an increasing trend in life expectancy for all the four countries. The gap between the two groups, Sweden and Norway on one side and Spain and Italy on the other keeps decreasing over time. In the latter four decades of the century, life expectancy improvements resulted from mortality reductions for younger ages and those over age 45. Notice that life expectancy in the 80s is almost the same for all the four countries and beyond that we can observe that women in Spain and Italy seem to have a longer life expectancy than women in Sweden and Norway. For the period 2014 the life expectancy at birth for women in Spain was 85.62 years, 85.17 years in Italy, 84.09 years in Norway and 84.05 years in Sweden. Thus, women in the Mediterranean are expected to live about one and a half year longer than those in Scandinavia. Progress in the treatment of cardiovascular disease and some forms of cancer on the one hand, and on the other the progress made in the prevention of certain "man-made" diseases, such at alcoholism, smoking or accidents, have made the life expectancy increase rapidly in the Mediterranean. (Vallin and Meslé 2004)

4.2 Cohort data

In order to have complete information on a cohort, it has to be observed from birth to extinction (i.e., the date by which all cohort members are assumed to have died). However, for our data this is only the case for the cohorts born around 1910. For the younger cohorts, we only have mortality information up to the age of the cohort in 2014.

4.2.1 Cohort mortality

Figure 4.6 shows the mortality rates of the cohorts born in 1910. We can observe that mortality gradually decrease for all the countries and reaches a minimum level around 12 years. Between the ages 0-8, we can see that Spain has the highest mortality, followed by Italy while the mortality in Sweden and Norway are almost the same. We observe a reduction in mortality from the age 15 until the age 50 for all those countries. This is most likely caused by increased living standard. From age 50 mortality starts to increase gradually and the gap between the countries decreases. On figure 4.6 we can observed that the mortality from age 70 is almost the same in all the four countries.

The mortality for the cohort born in 1930 is shown in figure 4.7. We can observe that there is a clear difference of mortality between the four countries from the age 5 years until the age 40 years. This is contrary to the cohort 1910 where the mortality in Sweden overlaps with Norway and the mortality in Spain with the one in Italy. Spain and Italy have the highest mortality up to ages 50-60 years, but it is the contrary for older ages.

Figure 4.8 shows that Norway (blue line) has the lowest level of mortality for the cohort 1950 with the minimum around the ages 10 years and 24 years. We also observe an unstable development in the Norwegian mortality with a lot of jumps from age 10 years to age 50 years. Her also the unstable development could be due to the very low mortality and the logarithmic scale. We also have low mortality in Sweden and Norway up to around 50 years, but we observe the contrary beyond that.

Figure 4.9 shows a decrease of the mortality for all the countries, reaching minimum mortality at age 12 years for Norway and 14 years for Sweden. Sweden and Norway show an increase from the minimum to a plateau for young adults. From age 15 to age 45, we can observe a quite stable and low variation of mortality for all the countries, forming a plateau.

We can conclude that the gap of mortality is bigger in the earlier age groups for cohort data than in period data. We also observed a slower decrease of mortality among the younger in the cohort data than in the period data. In the case of cohort, we can observe a smaller concentration of mortality in the earlier ages than in the older ages group. While Sweden and Norway have the lowest mortality in the younger ages, we observe the contrary for the aged.

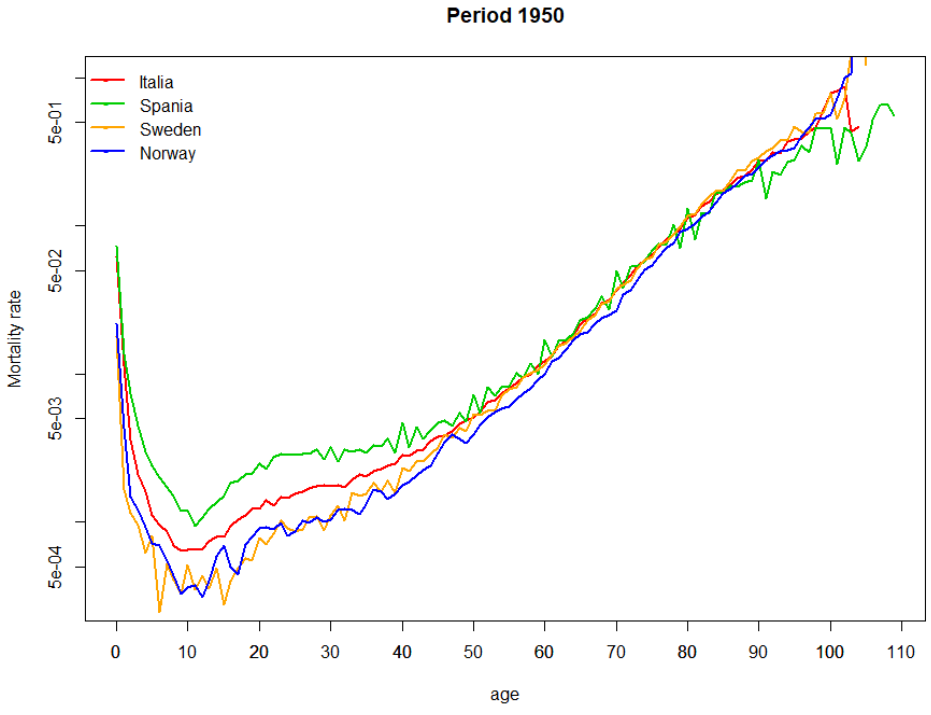


Figure 4.2: Period mortality rate for 1950 for all countries.

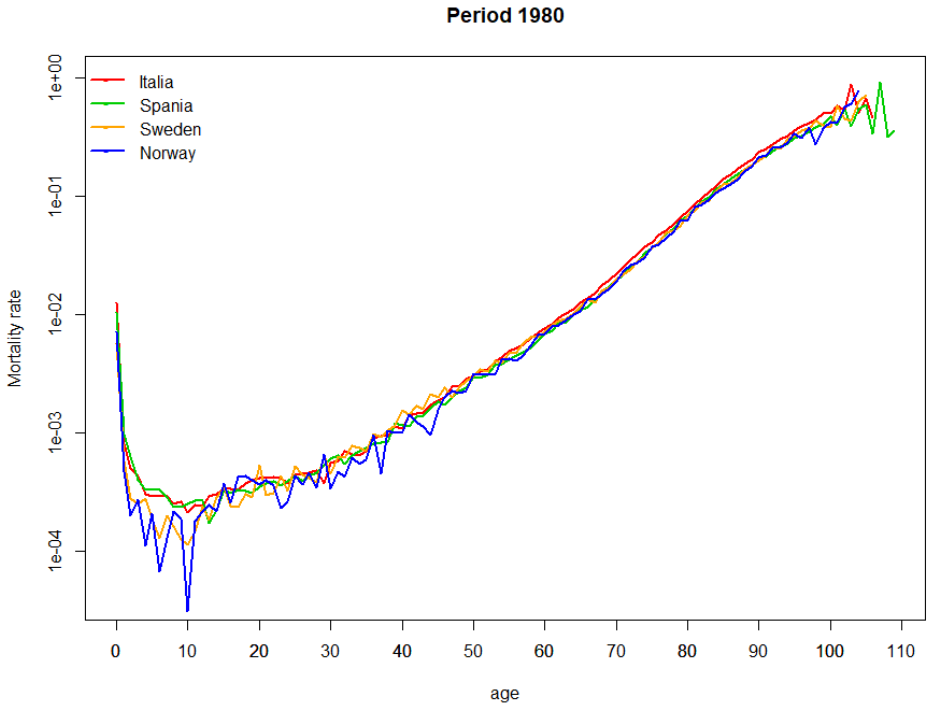


Figure 4.3: Period mortality rate for all countries.

4.2.2 Expected number of years lost for cohorts

In this section, we want to look at the expected number of year lost for cohort data. In order to compute the cohort life expectancy accurately, we need the complete mortality history of a cohort, and this is only possible for older cohorts that were born more than 100 years ago. For younger cohorts, we are not able to compute the cohort life expectancy. In this thesis, we therefore suggest that one may instead consider the expected number of years lost for a cohort up to a given age. (Andersen 2013) Then, by computing the expected number of years lost for a number of cohorts, one will obtain a good picture of the longevity in a country that is only based on the available data for the cohorts.

Figures 4.10 - 4.15 show the expected number of years lost up to the ages 90, 80, 70, 60, 50 and 40 in Spain, Italy, Sweden and Norway. The scales on the y-axis are adjusted in order to have a better visualization of the four countries. The expected number of years lost up to 90 years, figure 4.10, shows that Spanish (green plot) and Italian (red plot) women have lost more years of life than Swedish (orange plot) or Norwegian (blue plot). We can observe some little variations in Italy and Spain for the cohort 1919 and 1920 respectively.

The expected number of years lost up to age 80 in figure 4.11, have almost the same trend as in figure 4.10. Sweden and Norway remain close to each other but the gap between them and Spain and Italy is still very high. Here also we observe a little jump around the cohort 1930 but only in Spain.

Figure 4.12 shows the expected number of years lost up to age 70. We observe that the gap between all the countries is larger for 70 years than for 80 years. Women in Sweden and Norway lose fewer years than women in Spain and Italy.

On figures 4.14 and 4.13, we can observe that the expected number of years lost decreases with time for all the cohorts. But it seems to decrease faster in Spain and Italy than in Sweden and Norway. We can also observe an increase of variation for all the countries.

Figure 4.15 shows the plots of the expected number of years lost up to the age 50 for the four countries. Around the 70s we can see that Spain and Italy have gained about 12 and 14 years respectively but Sweden and Norway still have the lowest expected number of years lost.

Figures 4.16 - 4.19 show the expected number of years lost but this time we fix the cohort and look at the expected number of years lost as a function of age. The lines on these figures slope upwards to the right. As the cohort increases, the expected number of years lost decreases.

The gap between Spain and Italy on one side and Sweden and Norway on the other side has been considerably reduced as we can observe in figure 4.19. But for all cohorts and ages, women in Spain and

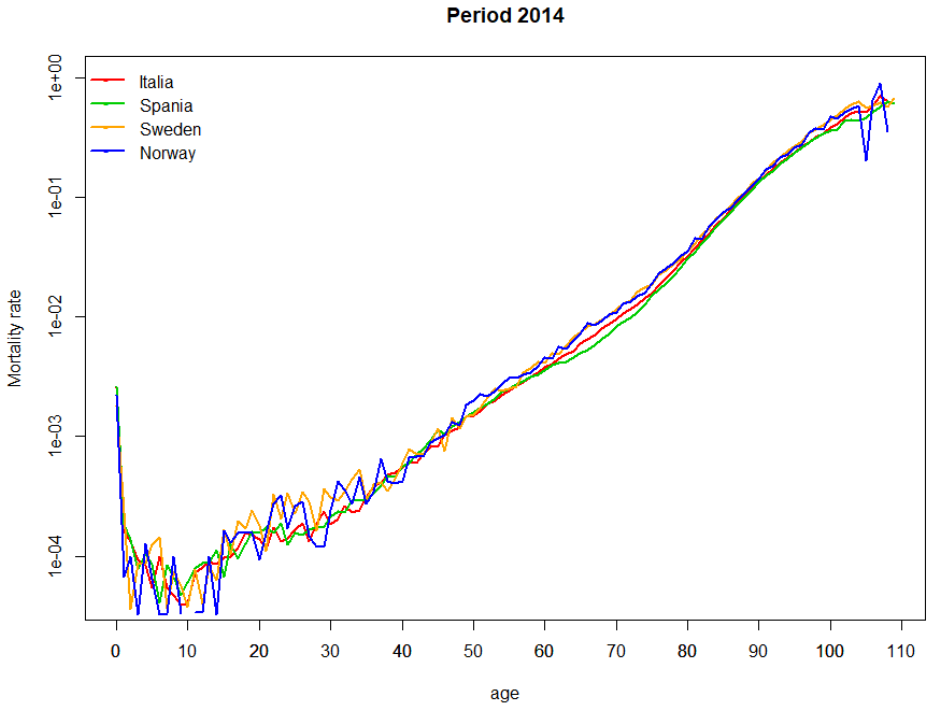


Figure 4.4: Period mortality rate for 2014 for all countries.

Italy may expect to loose more years than women in Sweden and Norway.

4.3 Summary of the chapter

The plots of our estimates suggest that during the 20th century, mortality rates have declined quite rapidly in the Mediterranean and the Scandinavian countries. We saw that mortality was highly concentrated among the younger and the aged in all those countries. Despite the fact that mortality was very high for all those countries, there was a huge gap between the two regions and that gap has continuously decreased over time. The steady reduction of mortality over the years leads to an improvement of live expectancy. The period data suggested that from the first decade of the 20th century to the 80s, women lived longer in Sweden and Norway than in Spain and Italy. But after the 80s we observed that women in Spain and Italy seemed to have caught up and even passed Sweden and Norway, having the higher life expectancy. But looking at the cohort data we can see a different picture. In fact the figures of the expected number of year lost tells us that women in Norway and Sweden are still expected to lose fewer years than those in Spain and Italy. In the next chapters we will look at a possible explanation of the conflicting results for period and cohort data.

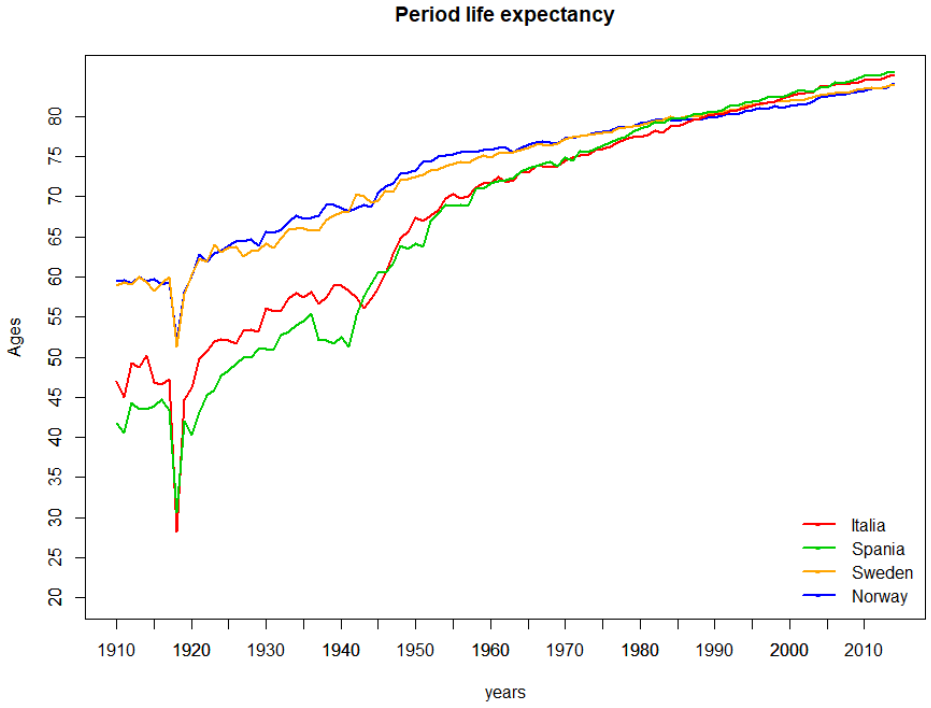


Figure 4.5: Period life expectancy for all the countries.

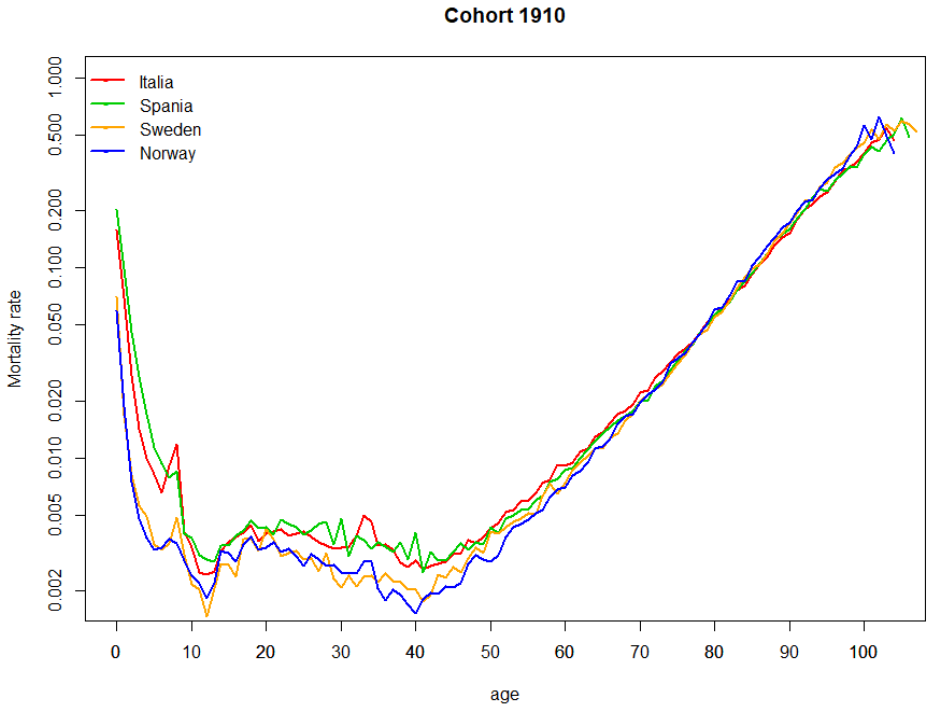


Figure 4.6: Cohort mortality rate for cohort born in 1910 for all countries.

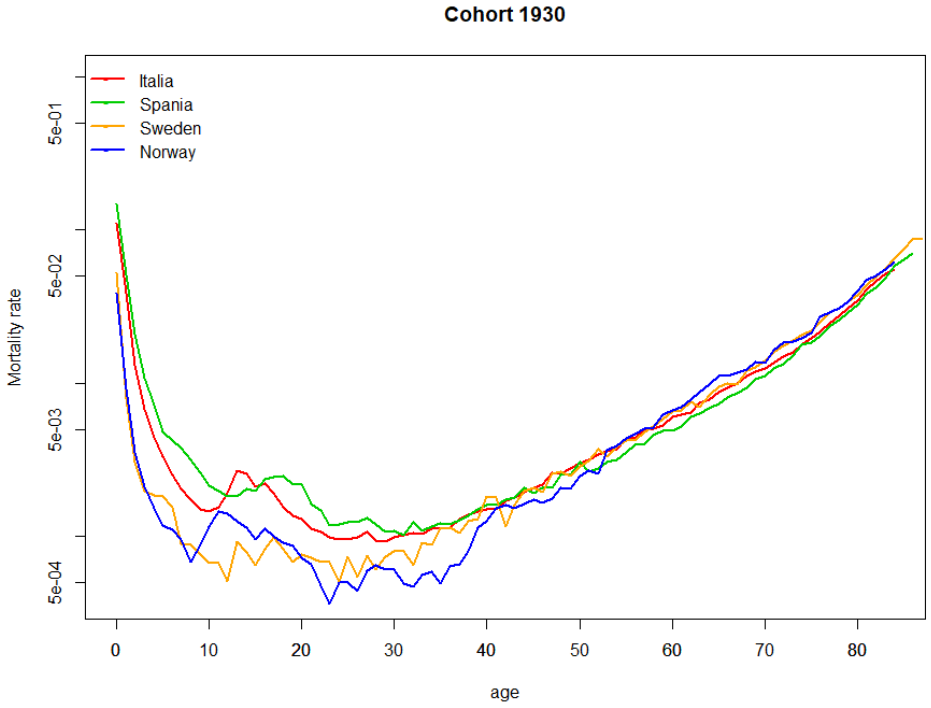


Figure 4.7: Cohort mortality rate for cohort born in 1930 for all countries.

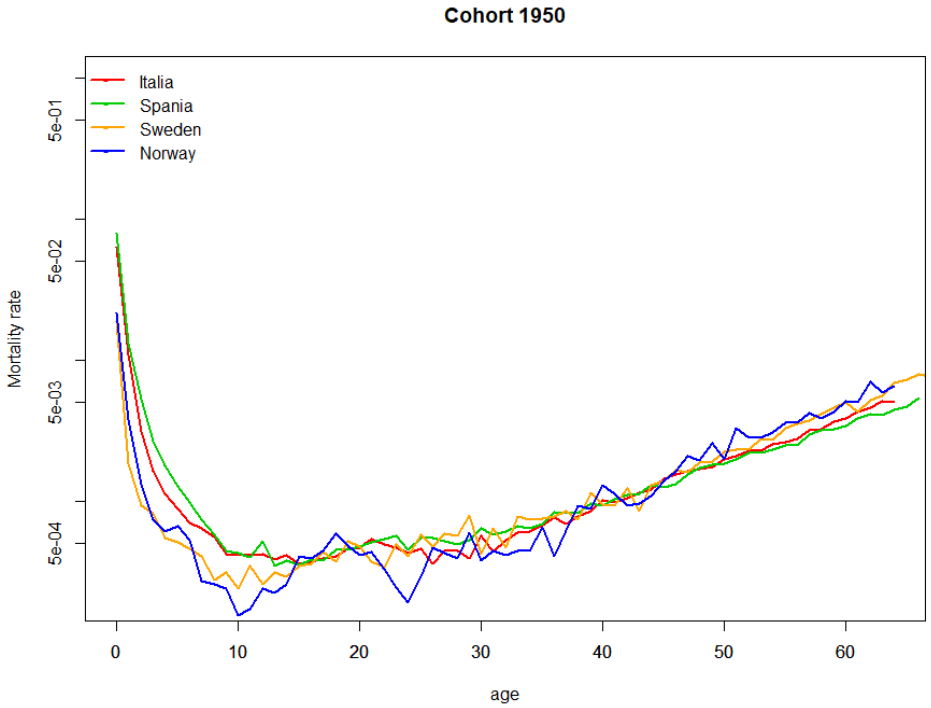


Figure 4.8: Cohort mortality rate for cohort born in 1950 for all countries.

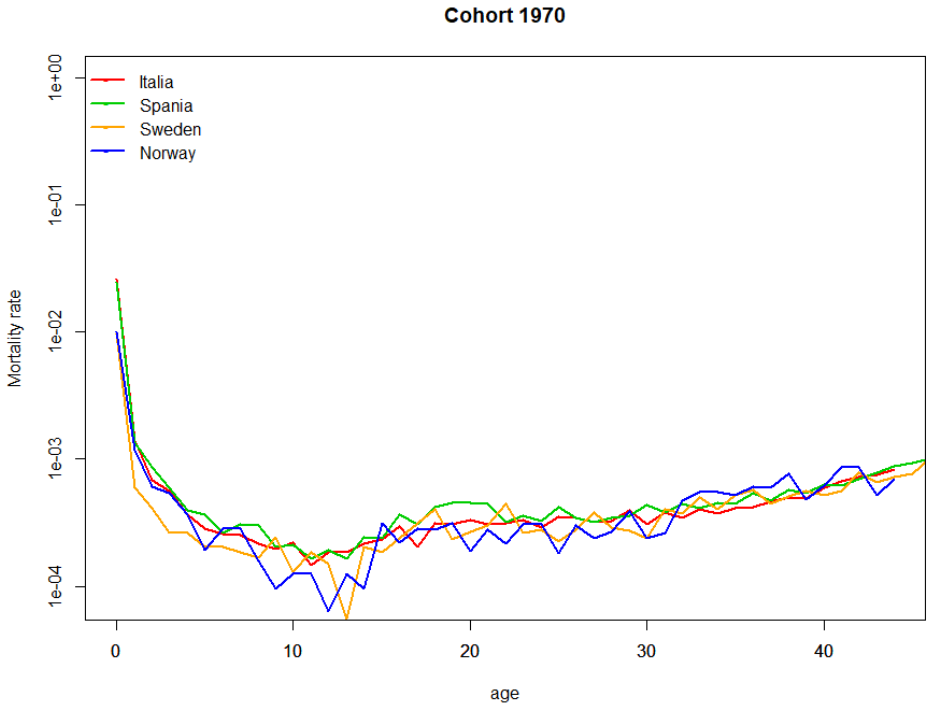


Figure 4.9: Cohort mortality rate for cohort born in 1970 for all countries.

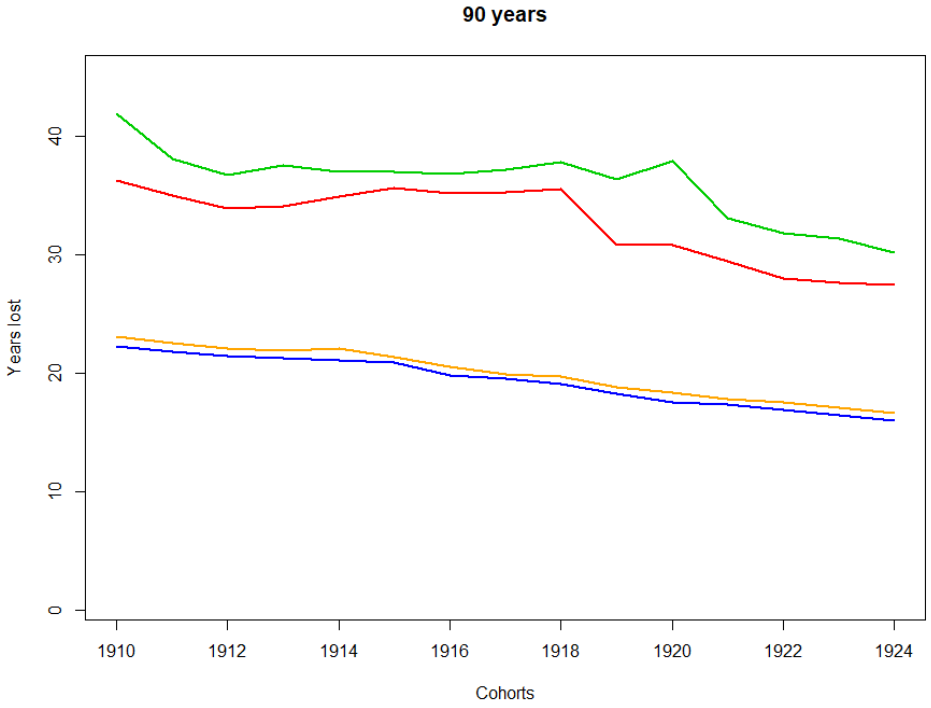


Figure 4.10: Number of years lost. Here we fix the age and consider the expected number of years lost as a function of cohort.

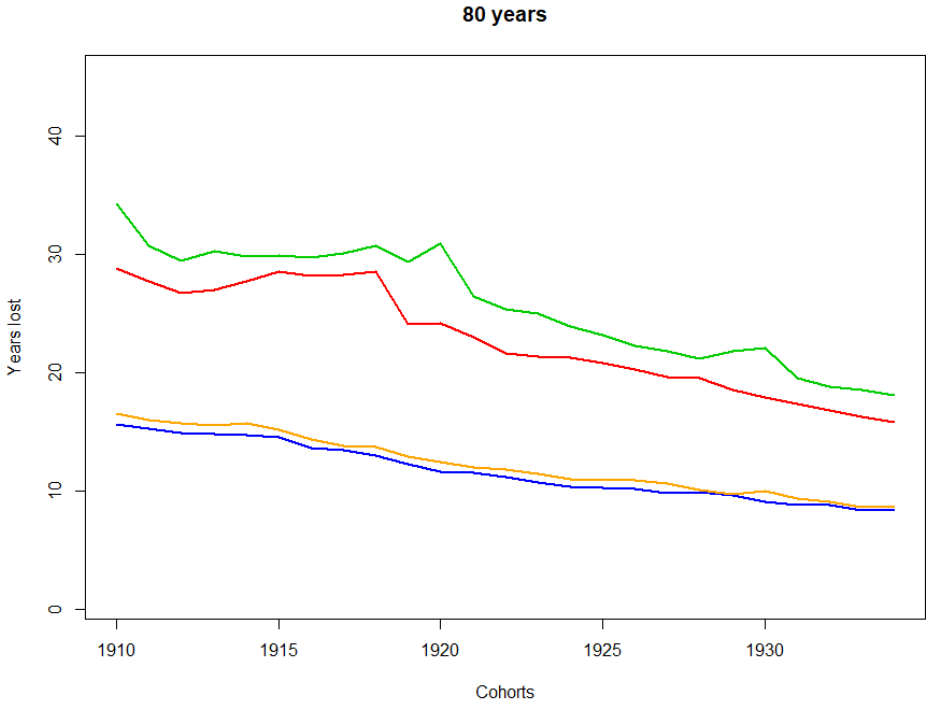


Figure 4.11: Number of years lost. Here we fix the age and consider the expected number of years lost as a function of cohort.

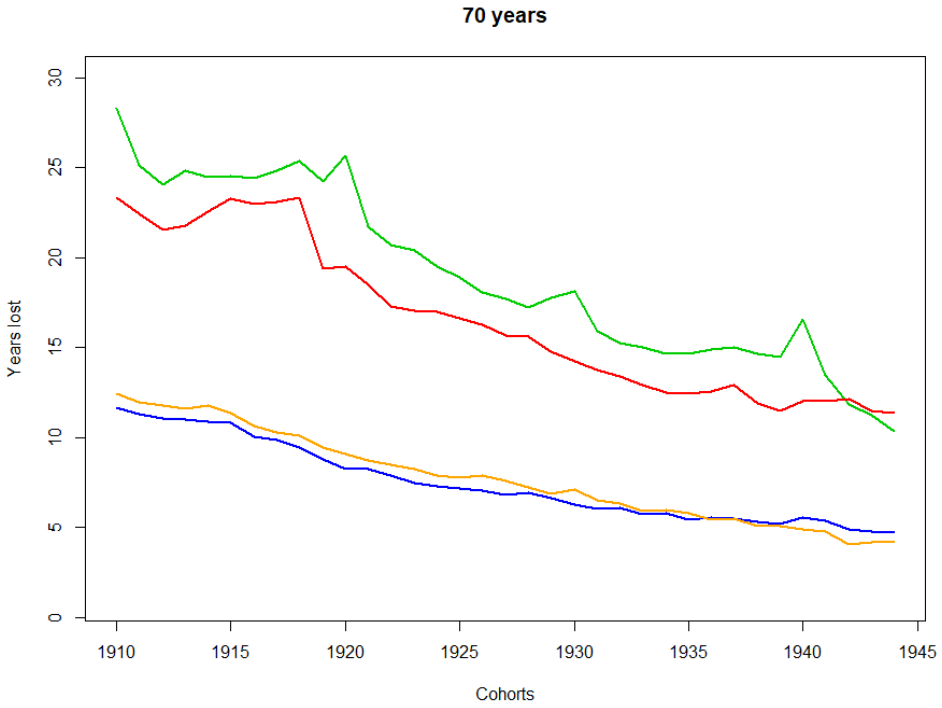


Figure 4.12: Number of years lost. Here we fix the age and consider the expected number of years lost as a function of cohort.

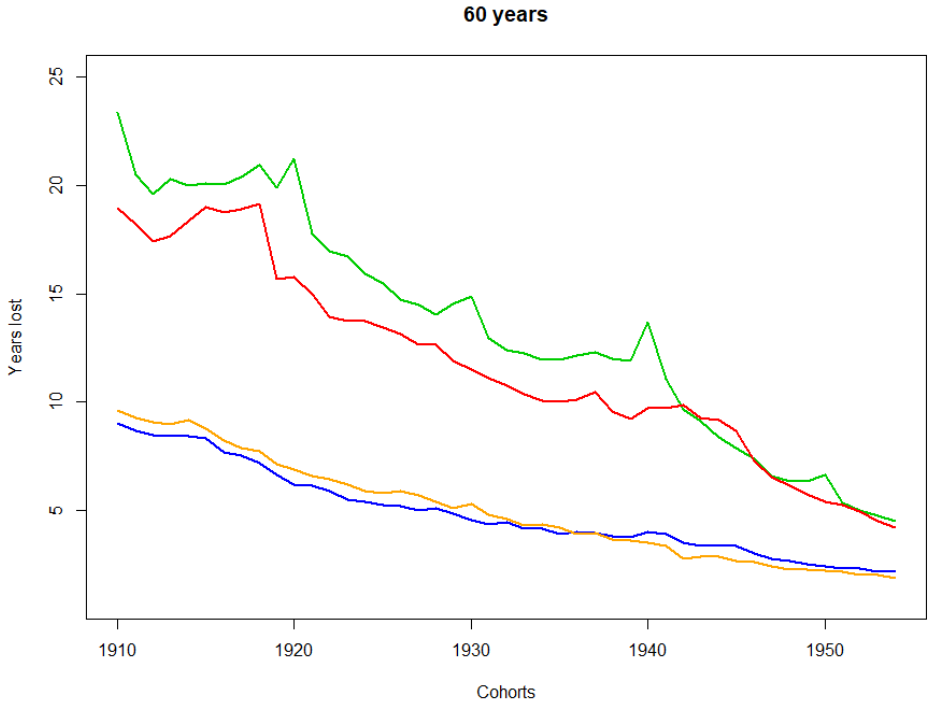


Figure 4.13: Number of years lost. Here we fix the age and consider the expected number of years lost as a function of cohort.

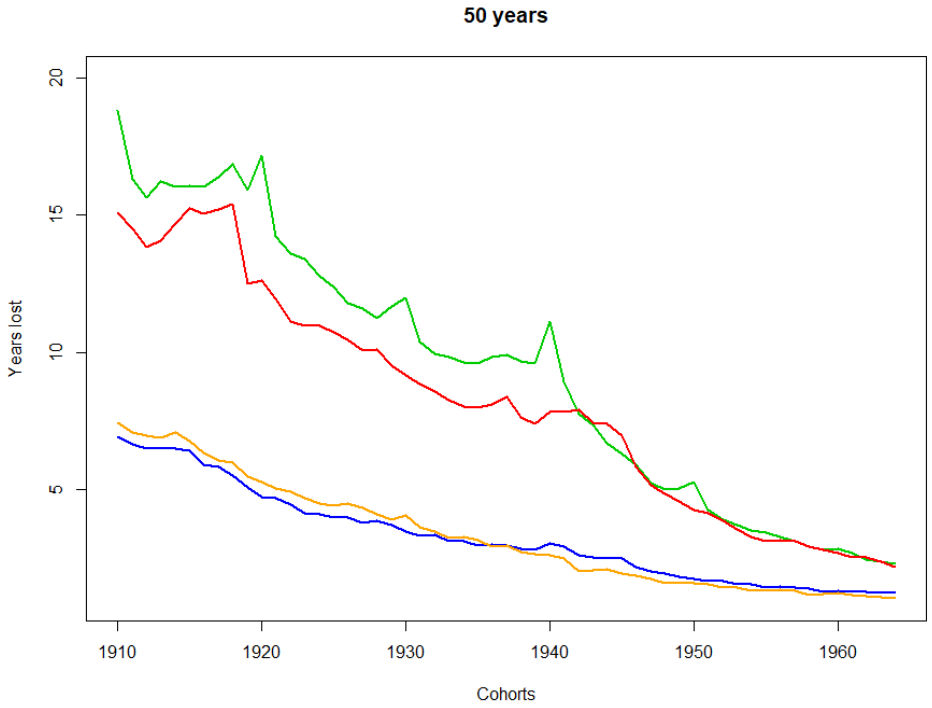


Figure 4.14: Number of years lost. Here we fix the age and consider the expected number of years lost as a function of cohort.

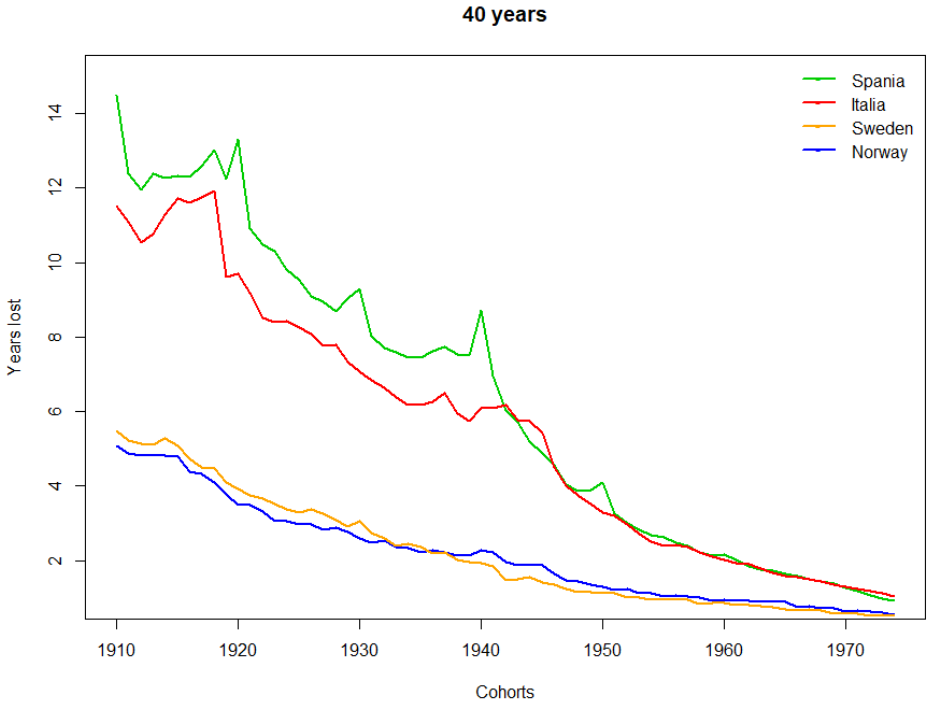


Figure 4.15: Number of years lost. Here we fix the age and consider the expected number of years lost as a function of cohort.

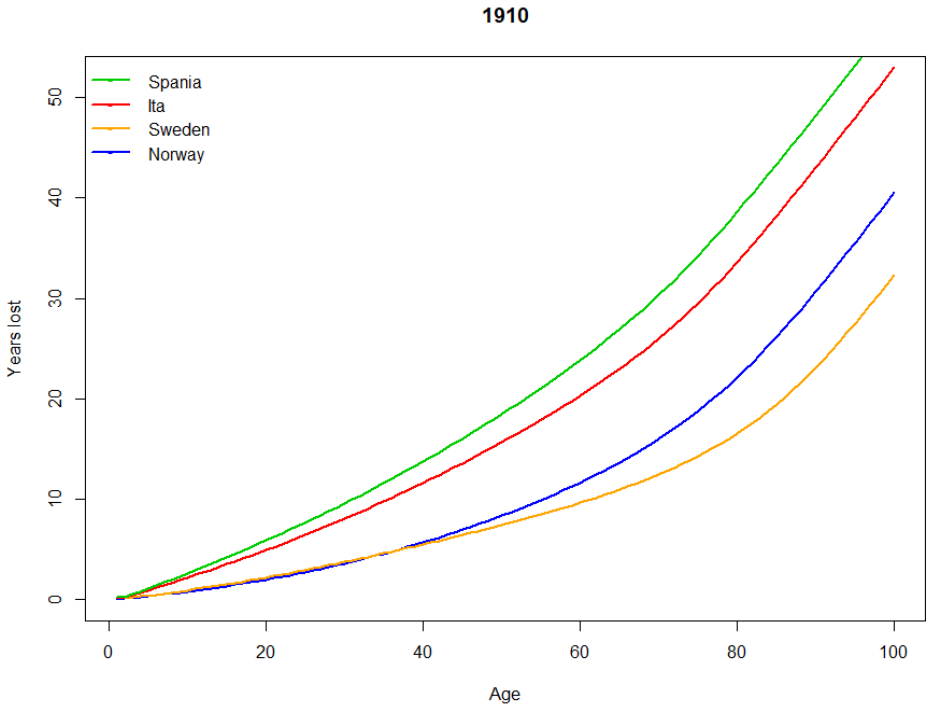


Figure 4.16: Number of years lost. Here we fix the cohort and look at the expected number of years lost as a function of age.

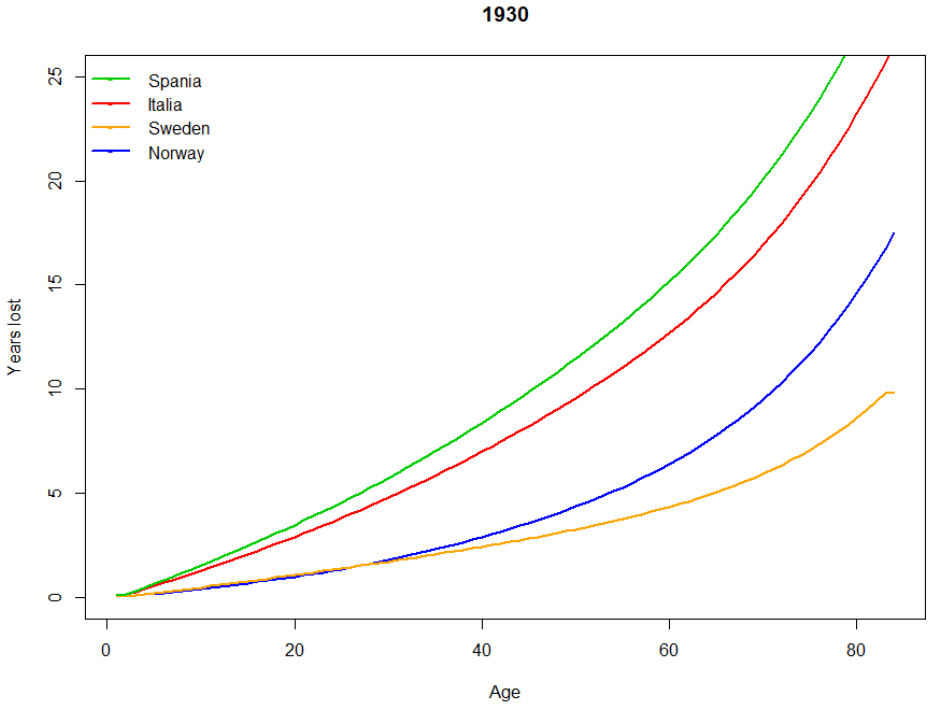


Figure 4.17: Number of years lost. Here we fix the cohort and look at the expected number of years lost as a function of age.

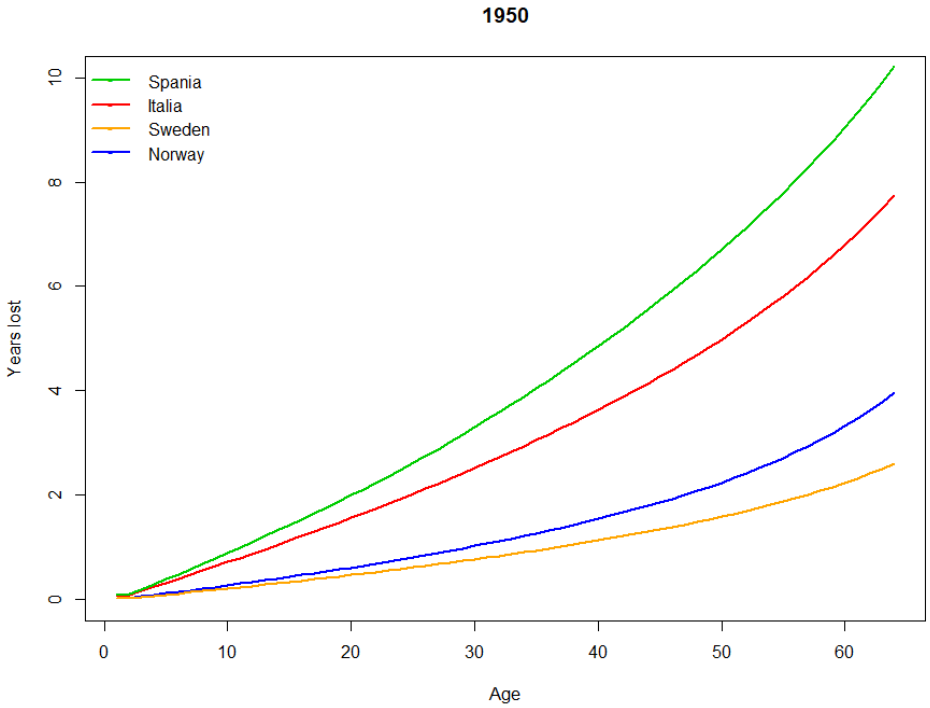


Figure 4.18: Number of years lost. Here we fix the cohort and look at the expected number of years lost as a function of age.

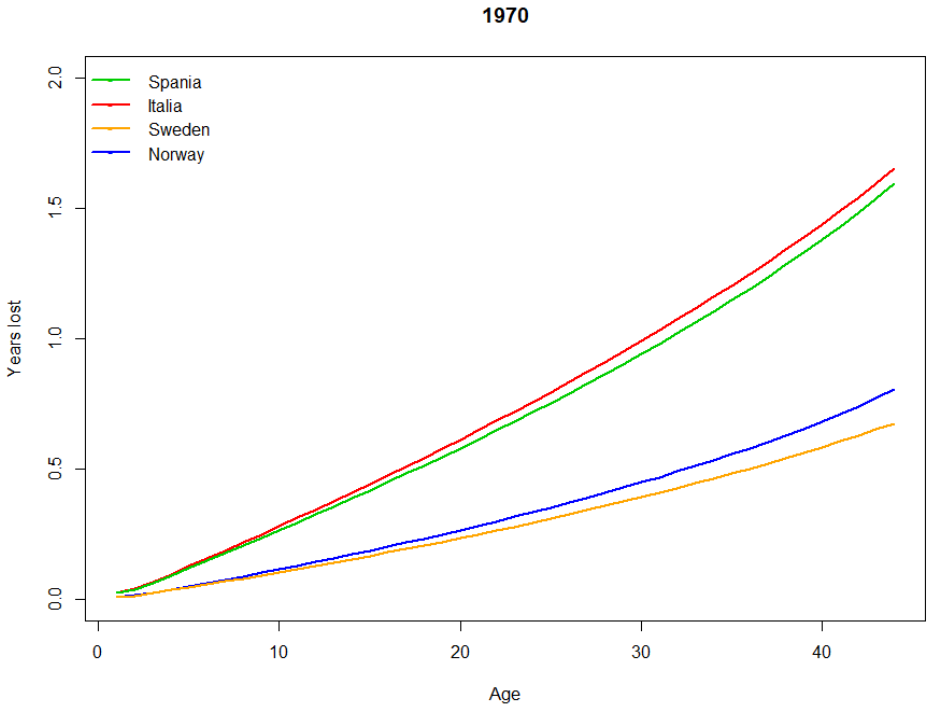


Figure 4.19: Number of years lost. Here we fix the cohort and look at the expected number of years lost as a function of age.

CHAPTER 5

FRAILTY MODELS

A frailty model is a random effects model for life times where the random effect (the frailty) has a multiplicative effect on the death intensity. It can be used to describe the influence of unobserved heterogeneity in a population. In a frailty model one has to distinguish between the individual death intensity and the population death intensity, where the individual death intensity refers to a single individual. Here the mortality varies among the individuals or groups of individuals due to specification of frailty for each groups. An individual with a frailty of 1 might be called a "standard" individual. As one may see from formula (5.2) below, if a standard individual has a 50 percent chance of surviving to some age, an individual with a frailty of 2 will have $(0.50)^2 = 0.25$, i.e 25 percent chance of surviving to this age, and an individual with frailty of $1/2$ on the other hand will have $(0.50)^{1/2} = 0.71$, i.e 71 percent chance of surviving to this age.

5.1 The frailty model

As in chapter 2, let the positive random variable T be the life time of an individual with corresponding c.d.f $F(t) = P(T \leq t)$ and p.d.f $f(t) = F'(t)$. Then the survival distribution is written as,

$$S(t) = 1 - F(t) = P(T > t)$$

and the instantaneous death rate or the hazard rate as,

$$\mu(t) = f(t)/S(t)$$

To get an alternative, and more directly interpretable expression for $\mu(t)$, note that

$$P(t < T \leq t + \Delta t | T > t) = [S(t) - S(t + \Delta t)]/S(t).$$

Hence using the survival distribution and the death intensity, we get

$$\mu(t) = -S'(t)/S(t) = \lim_{\Delta t \to 0} P(t < T \leq t + \Delta t | T > t)/\Delta t$$

So far we have assumed that the death intensity is the same for each individual. Now we will allow the death intensity to vary between individuals.

We assume that the death intensity of an individual is given as the product of an individual specific quantity Z and a basic rate $\alpha_0(t)$:

$$\alpha(t|Z) = Z \cdot \alpha_0(t) \tag{5.1}$$

Here Z is considered as a random variable over the population of individuals, specifying the level of frailty. The frailty Z captures heterogeneity in the population. Z and $\alpha_0(t)$ are unobservable. What may be observed in a population is not the individual death intensity, but the death intensity for the population. Frailties can therefore describe situations where what is observed on a population level may differ from what goes on at the individual level. The definition of the frailty assumes that in the population each individual comes to life with specific level of frailty and stays at this level all his or her life.

5.2 The population survival function and death intensity

Given the frailty model of one individual (5.1), we have the individual survival function:

$$S(t|Z) = P(T > t|Z) = \exp(-Z \cdot A_0(t)) \tag{5.2}$$

where $A_0(t) = \int_0^t \alpha_0(u)du$. The population survival function is found by integrating over the distribution of Z, that is,

$$\begin{aligned} S(t) = P(T > t) &= E[I(T > t)] \\ &= EE[I(T > t)|Z] \\ &= E[P(T > t)|Z)] \\ &= E[S(t|Z)] \\ &= E[\exp(-Z \cdot A_0(t))] \end{aligned}$$

Let $M_Z(t)$ be the moment generating function of Z:

$$M_Z(t) = E[\exp(tZ)]$$

The Laplace transform is related to the moment generating function, and for a positive random variables Z with density $g(z)$ it is given by:

$$L_Z(t) = M_Z(-t) = E[\exp(-tZ)] = \int_0^\infty \exp(-tz)g(z)dz$$

The population survival function can then be written as follows:

$$S(t) = E[\exp(-A_0(t) \cdot Z)] = L_Z(A_0(t)) \tag{5.3}$$

Using (5.3) the population death intensity denoted by $\mu(t)$ may now be written,

$$\mu(t) = -\frac{S'(t)}{S(t)} = \frac{L'_Z(A_0(t))}{L_Z(A_0(t))}\alpha(t) \tag{5.4}$$

The difference between the individual death intensity and the population death intensity is determined by the factor $\frac{L'_Z(A_0(t))}{L_Z(A_0(t))}$ in (5.4). In general the population death intensity cannot be interpreted as giving information on individual development in risk. In the population, the individuals with high frailty values will have the tendency of dying first. That will lead to a decrease of the frailty of the whole population. In other words, the value of the frailty in a population will decrease as the population gets older.

5.3 Modeling the Gamma frailty distribution

In the frailty literature, it is quite common to assume that the frailty Z is gamma distributed. The gamma distribution is chosen because it is a flexible distribution that takes on a variety of shapes as shown in figure 5.1. Frailty cannot be negative and the gamma distribution is, along with the log-normal and Weibull distribution, one of the most used distribution to model variables that are necessarily positive. The density of the gamma distribution is given as:

$$g(z) = \frac{1}{\beta^\alpha \Gamma(\alpha)} z^{(\alpha-1)} \exp(-z/\beta) \tag{5.5}$$

where β is the scale parameter and α the shape parameter.

Figure 5.1 plots the shape of gamma p.d.f's for four values of α. When $\alpha = 1$ it is identical to the well know exponential distribution; When $\alpha = 2$ we observe a more bell-shaped form. The moment generating function of the gamma distribution is given as:

$$M_Z(t) = \frac{1}{(1 - \beta t)^\alpha}$$

Using the Laplace transform, we obtain:

$$L_Z(t) = M_Z(-t) = \frac{1}{(1 + \beta t)^\alpha} \tag{5.6}$$

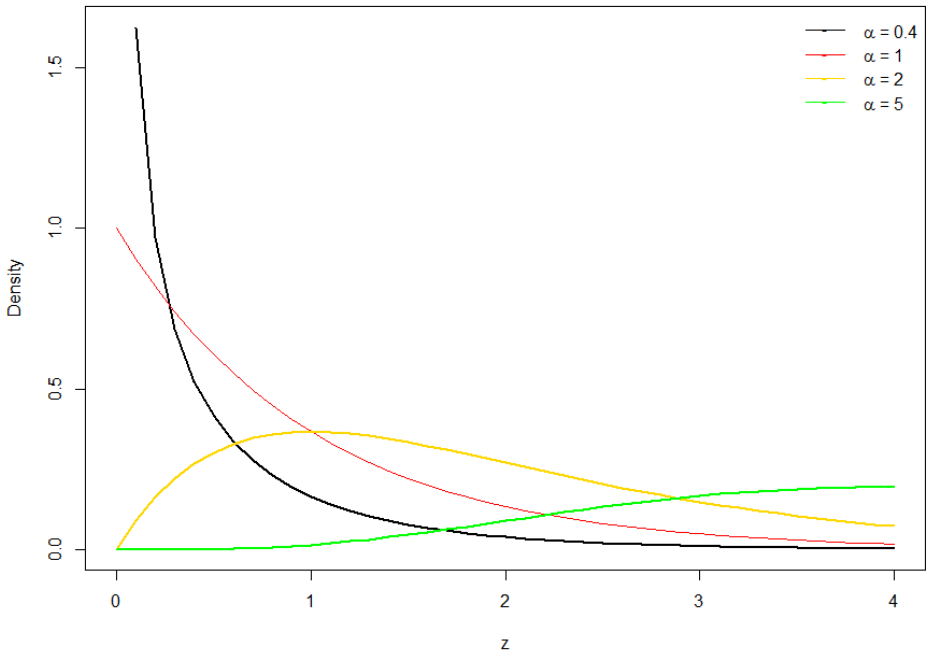


Figure 5.1: Gamma density for different values at α when $\beta = 1$.

It is common to assume that $E[Z] = 1$. It then follows that $\alpha\beta = 1$ and $\alpha = 1/\beta$. Further we then have that $V(Z) = \alpha\beta^2 = \beta$. Equation (5.6) then becomes:

$$L_Z(t) = M_Z(-t) = \frac{1}{(1+\beta t)^{1/\beta}} = (1+\beta t)^{-1/\beta} \tag{5.7}$$

Using (5.3), the survival function can now be written as:

$$S(t) = L_Z(A_0(t)) = (1+\beta A_0(t))^{-1/\beta} \tag{5.8}$$

Using (5.4), the population death intensity becomes:

$$\mu(t) = \frac{1}{1+\beta A_0(t)}\alpha_0(t) \tag{5.9}$$

This equation is useful because it gives a clear understanding of the effect of frailty on the death intensity of the population. It becomes clear to observe from (5.9) that when $\beta = 0$ there is no frailty and $\mu(t)$ and $\alpha_0(t)$ are identical. We observe that when the frailty of the population decreases, the death intensity increase. This can be observe in Figure 5.2 where we have chosen the basic rate $\alpha_0(t) = t^3$. We also observe that the population death intensity decreases with a strength determined by β.

5.4 Two points frailty distribution

An alternative to the gamma frailty model is a two-points model for the frailty. Here we assume that the frailty Z can take the two values z_1 and z_2 with probabilities $P(Z = z_1) = \pi_1$ and $P(Z = z_2) = \pi_2$; $\pi_1 + \pi_2 = 1$.
The Laplace transform becomes:

$$L_Z(t) = E[\exp(-tZ)] = \exp(-tz_1)\pi_1 + \exp(-tz_2)\pi_2 \tag{5.10}$$

The individual hazard can be written as :

$$\alpha(t|Z) = Z \cdot \alpha_0(t) \tag{5.11}$$

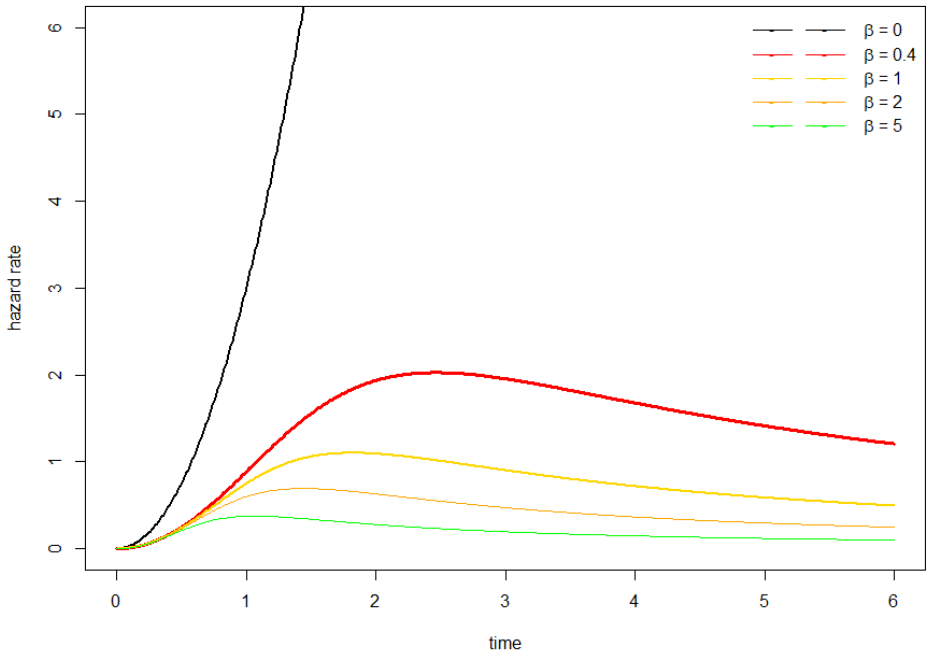


Figure 5.2: Population hazard rates with various values of β

With a two points frailty distribution we have two groups with respective death intensity, $z_1\alpha_0(t)$ and $z_2\alpha_0(t)$. The population survival function is given by

$$S(t) = L_Z(A_0(t)) = \exp(-A_0(t)z_1)\pi_1 + \exp(-A_0(t)z_2)\pi_2 \tag{5.12}$$

In order to derive the population death intensity, we differentiate the survival function

$$\begin{aligned} S'(t) &= \exp(-A_0(t)z_1)(-z_1\alpha_0(t))\pi_1 + \exp(-A_0(t)z_2)(-z_2\alpha_0(t))\pi_2 \\ &= -[\pi_1 z_1 \exp(-A_0(t)z_1) + \pi_2 z_2 \exp(-A_0(t)z_2)] \cdot \alpha_0(t) \end{aligned} \tag{5.13}$$

The population death intensity becomes :

$$\begin{aligned} \mu(t) &= -\frac{S'(t)}{S(t)} \\ &= \frac{\pi_1 z_1 \exp(-A_0(t)z_1) + \pi_2 z_2 \exp(-A_0(t)z_2)}{\pi_1 \exp(-A_0(t)z_1) + \pi_2 \exp(-A_0(t)z_2)} \cdot \alpha_0(t) \\ &= W(t)z_1\alpha_0(t) + (1 - W(t))z_2\alpha_0(t) \end{aligned} \tag{5.14}$$

where

$$W(t) = \frac{\pi_1 \exp(-A_0(t)z_1)}{\pi_1 \exp(-A_0(t)z_1) + \pi_2 \exp(-A_0(t)z_2)} \tag{5.15}$$

Thus the population death intensity is a weighted average of the death intensities in the two groups. The life expectancy up to age a is :

$$\begin{aligned} E[T_a] &= \int_0^a S(u)du \\ &= \pi_1 \int_0^a \exp(-A_0(u)z_1)du + \pi_2 \int_0^a \exp(-A_0(u)z_2)du \end{aligned} \tag{5.16}$$

If we introduce $S_0(t) = \exp(-A_0(t))$, we have

$$\exp(-A_0(t)z_i) = S_0(t)^{z_i}$$

for $i = 1, 2$. Thus we obtain

$$E[T_a] = \pi_1 \int_0^a S_0(u)^{z_1} du + \pi_2 \int_0^a S_0(u)^{z_2} du \tag{5.17}$$

5.5 Gompertz-Makeham

We now assume a Gompertz-Makeham model for the basic hazard with parametrization

$$\alpha_0(t) = a + ac^t = a + b\exp(t\log(c)) \tag{5.18}$$

where the constant a can be interpreted as the risk of death from all causes which do not depend on age and the second element of the equation is the age-dependent component, which increases exponentially with age. Then we have

$$\begin{aligned} A_0(t) &= \int_0^t \alpha_0(u)du \\ &= \int_0^t (a + b\exp(u\log(c)))du \\ &= \left[au + \frac{b}{\log(c)}\exp(u\log(c))\right]_0^t \\ &= at + \frac{b(\exp(t\log(c)) - 1)}{\log(c)} \\ &= at + \frac{b(c^t - 1)}{\log(c)} \end{aligned} \tag{5.19}$$

The corresponding survival function is:

$$S_0(t) = \exp(-A_0(t)) = \exp(-at - \frac{b(c^t - 1)}{\log(c)}) \tag{5.20}$$

The population hazard may then be obtained from (5.14) and (5.15) using (5.18) and (5.19) and the life expectancy up to age a is given by (5.17).

Figure 5.3 shows the baseline death intensity $\alpha_0(t)$ of the Gompertz-Makeham form (black line) and the population hazards for a two-points frailty distribution when $a = 5 * 10^{-4};\ b = 2 * 10^{-5};\ c = 1.10$ and the probabilities are $\pi_1 = 0.7$ and $\pi_2 = 0.3$.The red line in Figure 5.3 is the population hazard when the frailty values are $z_1 = 0.5$ and $z_2 = 2$. We observe that the black and the red line overlap and are steadily increasing until the age of 70 years where μ (plot in red) starts slowing down. The population life expectancy is here 83.95 years. We then increase the value of the frailty z_2 from 2 to 3 while z_1 remains the same and we obtain a new population hazard $\mu(t)$ (plot in blue). We observe on the new plot of the population hazard that at younger ages the death intensity is higher than the basic hazard $\alpha_0(t)$ and also higher than the population hazard (plot in red) with the lower value of z_2. But from age 60 we observe a change of trend, the population hazard rate μ that had the highest mortality at the younger ages (blue line) increases less until age 80 where it cross the red line. With the new frailty value, the population life expectancy becomes 82.39 years. That is the phenomena that was observed in chapter 4 where in Figures 4.6 - 4.8 the cohorts of Norwegian and Swedish women had a lower mortality at the younger ages than the women in Spain and Italy, but the opposite was the case for older ages. We can conclude that the higher mortality observed in the population hazard $\mu(t)$ (plot in blue) for younger ages is the result of a higher number of individuals with a big value of frailty, since those individuals have a higher chance of dying than those with a smaller value of frailty. As the population get older, we have diminution of the number of frail individuals leading to the reduction of the mortality. That is why population hazard $\mu(t)$ (plot in blue), despite a higher mortality among the younger, has the lowest mortality among the older. In the next chapter, we will use the Gompertz-Makeham frailty distribution model to study the relationship between the period and the cohort mortality.

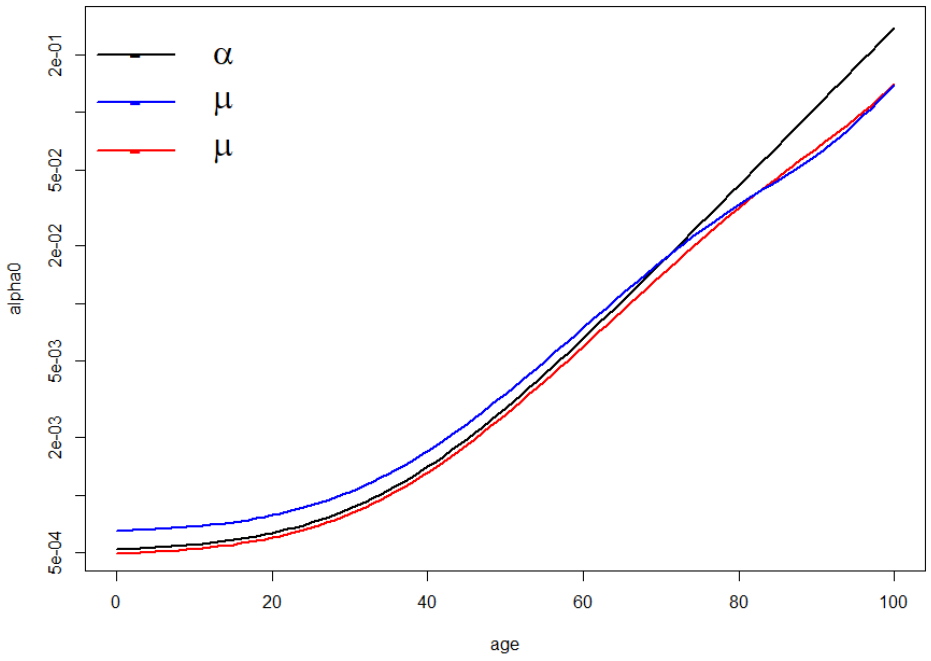


Figure 5.3: Basic hazard rate $\alpha_0(t)$ and population hazard rates $\mu(t)$ for the two points Gompertz-Makeham frailty distribution. See text for details

CHAPTER 6

COHORT AND PERIOD LIFE EXPECTANCY USING THE TWO-POINTS GOMPERTZ-MAKEHAM FRAILTY DISTRIBUTION

We observed in chapter 4 that in the early part of the 20th century, there was a very large difference in life expectancy between Spain and Italy on one hand and Norway and Sweden on the other. Over the last decade, that inequality has decreased and the period data suggest that since the 80's women in Spain and Italy are living longer than those in Norway and Sweden. See Figure 4.5. But the cohort data suggest that women in Scandinavia are still living longer. see Figures 4.10 - 4.19. In this chapter we will illustrate with the help of the two-point Gompertz-Makeham frailty model that women in Italy and Spain living longer than those in Norway and Sweden may just be an artifact.

6.1 Two-points frailty distribution for cohort

In the following, we will study hundred cohorts over a period of hundred years. We index the cohort by $k = 1; \ldots ; 100$. We assume a two-points frailty distribution for cohort k with a baseline death intensity of the Gompertz-Makeham form

$$\alpha_{0k}(t) = a_k + b_k c_k^t \tag{6.1}$$

Assume that the frailty Z_k for cohort k can take two values z_{1k} and z_{2k} with probabilities $P(Z = z_{1k}) = \pi_{1k}$ and $P(Z = z_{2k}) = \pi_{2k}$; $\pi_{1k} + \pi_{2k} = 1$. For the cohort k it follows from (6.1) that:

$$A_{0k}(t) = \int_0^t \alpha_{0k}(u)du = a_k t + \frac{b_k(c_k^t - 1)}{\log(c_k)}$$

The population survival function for the cohort k is given by:

$$S_k(t) = \exp(-A_0(t)z_{1k})\pi_{1k} + \exp(-A_0(t)z_{2k})\pi_{2k} \tag{6.2}$$

cf. (5.12) The corresponding population hazard is obtain from (5.14) and (5.15)

$$\mu_k(t) = W_k(t)z_{1k}\alpha_{0k}(t) + (1 - W_k(t))z_{2k}\alpha_{0k}(t) \tag{6.3}$$

where

$$W_k(t) = \frac{\pi_{1k}\exp(-A_{0k}(t)z_{1k})}{\pi_{1k}\exp(-A_{0k}(t)z_{1k}) + \pi_{2k}\exp(-A_{0k}(t)z_{2k})} \tag{6.4}$$

Using that $S_{0k}(t) = \exp(-A_{0k}(t))$, the expected life length of the cohort k up to a given age a for two-point Gompertz-Makeham frailty model is then:

$$E[T_a] = \pi_{1k}\int_0^a S_{0k}(u)^{z_{1k}}du + \pi_{2k}\int_0^a S_{0k}(u)^{z_{2k}}du \tag{6.5}$$

cf. (5.17)

6.2 Two-points frailty distribution for period

In order to compute period mortality for a period of 100 years, we need information that describe how the parameters of the two-point Gompertz-Makeham frailty model have (hypothetically!) developed over a period of 200 years (namely the 100 years we focus on and a period of 100 years before that). We index these 200 years by $p = -99, -98, \ldots, 0, \ldots, 100$. As discussed in chapter 3 the period mortality for period p and age t corresponds to the cohort mortality at age t for the cohort born in year $k = p - t$. The period population hazard can therefore be obtain by:

$$\mu^{(p)}(t) = \mu_{p-t}(t) \tag{6.6}$$

where p is the period and $t = 0, 1, 2, \cdots, 100$ is the age. Because of the difficulty to integrate the hazard function, an approximation can be given by:

$$\int_0^t \mu^{(p)}(u)du \approx \sum_{x=0}^{t-1} \mu^{(p)}(t) \tag{6.7}$$

The population survival function of the period is obtain by:

$$S^{(p)}(t) = \exp(-\int_0^t \mu^{(p)}(u)du) \tag{6.8}$$

and the life expectancy of the period can be obtain by:

$$E^{(p)}[T] = \int_0^{\omega} S^{(p)}(t)dt \approx \sum_{x=0}^{\omega-1} S^{(p)}(t) \tag{6.9}$$

where ω is the maximum possible age.

6.3 Comparison of the period and cohort mortality using the two-points frailty distribution with the baseline death intensity of the Gompertz-Makeham

The parameters choosen for the various cohorts in the two-points Gompertz-Makeham frailty model are described in table 6.1. Here country A corresponds to Norway and Sweden and country B to Italy and Spain. Figure 6.1 compares the population mortality for country A and B for three different cohorts ($k = 1, 50, 100$). Cohort 1 is the oldest cohort and cohort 100 is the youngest cohort. The rates are on logarithmic scale.

For cohort 1, we observe a huge difference of mortality rates between country A and country B among the younger ages. That is due to higher number of frail individuals in country B than in country A. The mortality rates are therefore lower in country A than in country B. After the age of 60 years, we observe that the number of frail individuals have steadily decreased resulting in country B having lower mortality rates than country A.

For cohort 50, we observe that the mortality rates are greater in country B than in country A from birth until the age of 45 years. Beyond the age of 45 years, country A and country B have about the same mortality rates (solid and dotted red lines). The overlap of mortality in country A and country B after the age of 45 year is mainly due to the fact that frail individuals in country B will tend to die first and those remaining will have the same frailty level as individuals in country A.

Finally we observe that in cohort 100, there is no difference in mortality in country A and country B (solid and dotted blue lines).

Figure 6.2 shows the cohort life expectancy in country A in blue and country B in red. We observe that country A has the highest life expectancy back in time. But for the younger cohorts the life expectancy in country B increases and attains the same level of life expectancy as in country A.

In figure 6.3 we have an illustration of the period life expectancy in country A and country B. We observe that country A has the highest life expectancy during the earlier period. But the improvement of mortality in country B result in a rapid increase of life expectancy. This will result in the cross over we observed in the figure, where country B have higher life expectancy than country A for the most recent periods.

Based on the empirical study we had in the earlier chapter and the illustrative result we observed in this chapter, we can conclude that what we observed on Figure 4.5 in chapter 4 may be due to a selection effect and may therefore be an artifact. Since the 18th century, life expectancy have been higher in Norway

Table 6.1: Values of the parameters in the two-points Gompertz-Makeham frailty model for 200 cohorts (indexed as $k = -99, -98, \cdots, 0, 1, 2, \cdots, 100$)

Parameters	Country A	Country B
a	Equal to 0.005 for all $k = -99, -98, \cdots, 0$. Then decreasing exponentially to 3.10^{-5} for $k = 100$	Equal to 0.010 for all $k = -99, -98, \cdots, 0$. Then decreasing exponentially to 3.10^{-5} for $k = 100$
b	Equal to 2.10^{-5} for all $k = -99, -98, \cdots, 0$. Then decreasing exponentially to 10^{-5} for $k = 100$	Same as for country A.
c	Equal to 1.100 for all $k = -99, -98, \cdots, 100$.	Equal to 1.098 for all $k = -99, -98, \cdots, 0$. Then increasing exponentially to 1.100 for $k = 100$
π_1	Equal to 0.50 for all $k = -99, -98, \cdots, 100$.	Same as for country A.
z_1	Equal to 0.50 for all $k = -99, -98, \cdots, 100$.	Equal to 0.40 for all $k = -99, -98, \cdots, 0$. Then increasing linearly to 0.50 for $k = 50$ and staying at that value until $k = 100$
z_2	Equal to 2 for all $k = -99, -98, \cdots, 100$.	Equal to 3 for all $k = -99, -98, \cdots, 0$. Then decreasing linearly to 2 for $k = 50$ and staying at that value until $k = 100$

and Sweden than Italy and Spain, and that trend still continue despite the significant reduction of the expected life length gap between countries in the Mediterranean and those in Scandinavia.

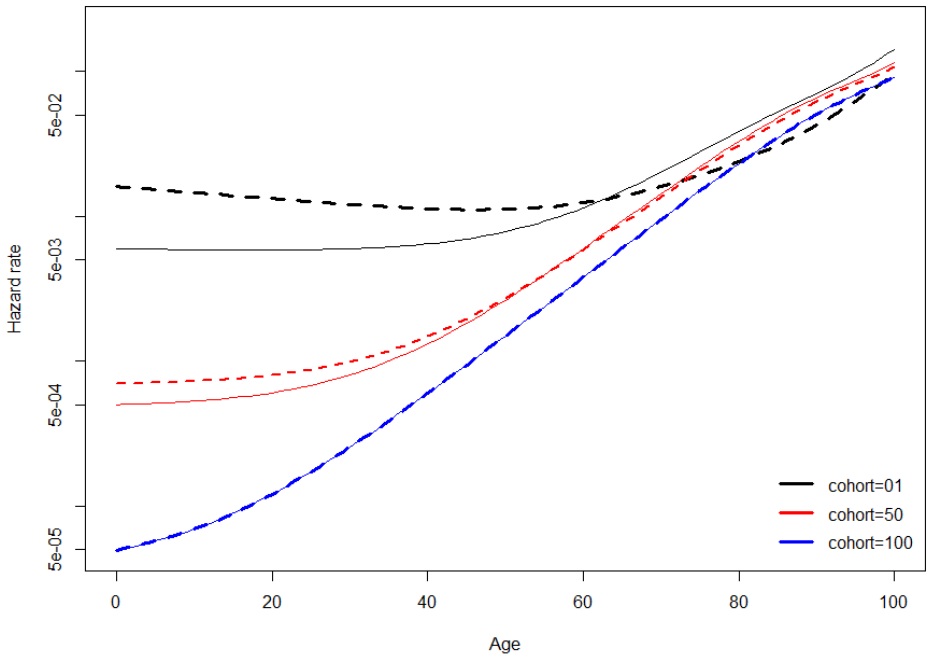


Figure 6.1: Population mortality for the two points Gompertz - Makeham frailty model for three cohorts (k= 01, 50, 100) for country A (drawn line) and country B (dashed line)

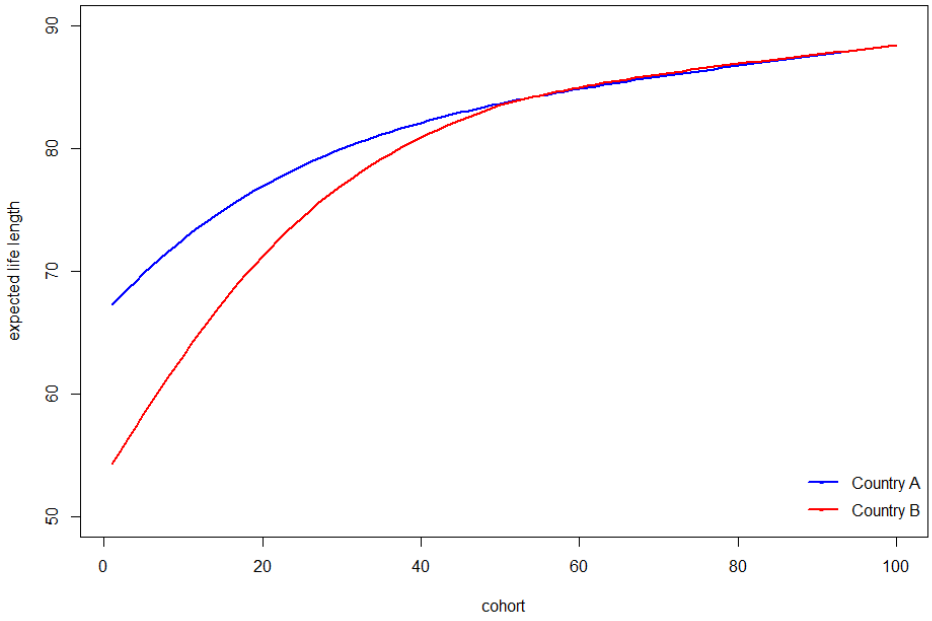


Figure 6.2: Cohort life expectancy illustrating country A and country B.

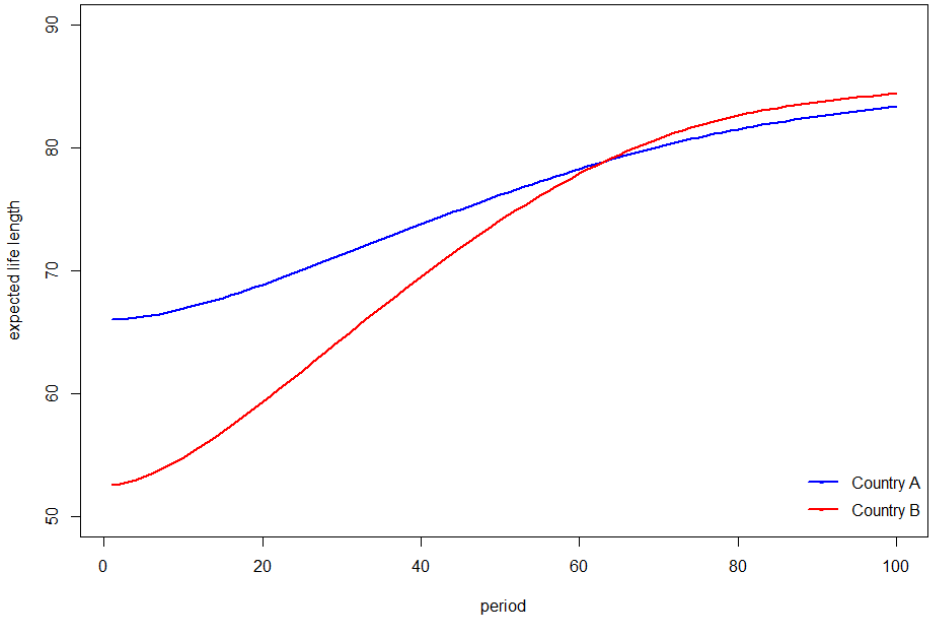


Figure 6.3: Period life expectancy illustrating country A and country B.

CHAPTER 7

CONCLUSION AND DISCUSSION

The aim of this study was to interpret and model the changes in life expectancy in Northern Europe and the Mediterranean countries. This was done in a few chapters. We started in the second chapter by considering some basic concepts and results in survival analysis and also considered parametric inference for survival data.

In the third chapter we introduced the Lexis diagram and how age, life line, period and cohort data are visualized in the diagram. We then defined and computed mortality rates for Norwegian females from the calendar year 1900 until 2014 for period data. Using those mortality rates we calculated the life expectancy at birth for some few years and observed that it has considerably increased over the past hundred years, going from 55.09 years in 1900 to 84.09 years in 2014. (See Figure 3.13). We also computed the life expectancy up to a certain age for different cohorts and noticed some differences with what was obtained in the case of period data. For example, the life expectancy up to age 100 years for the cohort born in 1900 was 62.43 years. Because of the non-availability of information for the cohorts data, we computed the expected number of years lost up to a certain age for some cohorts and found out that is has decreased considerably with time.

In order to have a good representation of the historical mortality development in North-Europe and the Mediterranean, we chose in chapter 4 to focus on the life expectancy for women in Spain, Italy, Norway and Sweden. Since Spain is not available in the Human Mortality Database before the year 1908, we used data from the year 1910 to 2014 such that we could have uniform data for all the four countries. Figure 4.5 indicated that Norway and Sweden had longest life expectancy during 1910-1970, but because of an improvement of the economical and life condition in the Mediterranean, Spain and Italy became the countries with highest life expectancy from the 80s. Based on this analysis, one could conclude that women in the Mediterranean are expected to live longer than women in Scandinavia.

But these results for period data correspond to an hypothetical situation, in real life women in a specific country are born in a cohort and live their lives years after years under the changing living conditions of that country. Therefore it would be accurate to compare the countries based on the cohort data. The difficulty with this approach is the non-availability of the data. To use data for a cohort, we need to know the complete mortality history of that cohort, i.e all the individuals of that cohort have to be dead. We therefore chose to compute the expected number of years lost up to a maximum age for a specific cohort. Figures 4.10, ..., 4.15 illustrated the expected number of years lost up to ages 90, 80, 70, 60, 50 and 40 as a function of their respective cohorts. We observed that the expected number of years lost was larger in Spain and Italy for the cohort 1910. For the youngest cohorts, we observed a reduction of the gap between the two regions. Despite the considerable reduction of the expected number of years lost in the countries from the Mediterranean, all the cohorts indicated that women in Norway and Sweden are still expected to lose less years than those in Spain and Italy.

To explain the difference of longevity observed in the case of period and cohort data, we introduce in chapter 5 and chapter 6 a two-points frailty distribution with the baseline death intensity of the Gompertz-Makeham form. To illustrate our analysis, we grouped the countries as follow: Norway and Sweden representing group A, Spain and Italy representing group B. We used information contained in table 6.1 to plot Figures 6.1,6.2, 6.3. Figure 6.1 indicated that country B had a higher mortality at younger ages than country A. The higher mortality in country B being the result of a big number of frail individuals. The mortality in country B get relatively lower and lower with age, caused by a reduction of the average number of frail individuals. The remaining individuals in country B had a lower frailty level leading to reduction of the population hazard rates. That may be the reason country B cross over country A in the case of period data as observed in figure 6.3, becoming the country with the higher life expectancy at the older ages. In the case of cohort, despite the reduction of the number of frail individuals

in country B, we observed in figure 6.2 that the life expectancy in country B increased with age but do not cross country A.

Hence if we consider life expectancy based on period mortality in figure 6.3, we may conclude that women in country B have higher life expectancy than those in country A. However if we consider the expected number of years lost for different cohorts in country A and B, we observed that women in country A are still expected to live longer than those in country B. Naive acceptance of observed population patterns may lead to erroneous policy recommendations if an intervention depends on the response of individuals. Furthermore, because patterns at the individuals level may be simpler than composite population patterns, both theoretical and empirical research may be unnecessarily complicated by failure to recognize the effects of heterogeneity. (Vaupel and Yashin 1985)

Bibliography

Aalen, O. O., Borgan, Ø., and Gjessing, H. K. (2008). *Survival and Event History Analysis. A Process Point of View*. Statistics for Biology and Health. Springer, New York.

Andersen, P. (2013). "Decomposition of Number of Life years lost according to causes of death". In: *Statistics in Medicine* 32, pp. 5278–5285.

Borgan, Ø. (1990). *Lectures Notes in Life History Analysis*. Vol. 01/90. Statistical Memoire. Institute of Mathematics University of Oslo.

Borgan, Ø. and Keilman, N. (2019). "Do Japanese and Italian Women Live Longer than Women in Scandinavia?" In: *European Journal of Population* 35, pp. 87–99.

Carstensen, B. (2007). "Age–period–cohort models for the Lexis diagram". In: *Statistics in Medecine* 26, pp. 3018–3045.

Clark, T. G. et al. (2003). "Survival Analysis Part I: Basic concepts and first analyses". In: *British Journal of Cancer* 89, pp. 232–238.

Sønstebø, A. (2019). "Andelen som dør, har aldri vært mindre". In: *Statistikk sentralbyrå, URL:www.ssb.no/befolkning/artikler-og-publikasjoner/andelen-som-dor-har-aldri-vaert-mindre*.

Vallin, J. and Meslé, F. (2004). "Convergences and divergences in mortality: A new approach of health transition". In: *Demographic Research* 2, pp. 11–44.

Vandeschrick, C. (2001). "The Lexis diagram, a misnomer". In: *Demographic Research* 4, pp. 97–124.

Vaupel, J. W. and Yashin, A. I. (1985). "Heterogeneity's Ruses: Some Surprising Effects of Selection on Population Dynamics". In: *The American Statistician* 39, pp. 176–185.

Wilmoth, J. R. et al. (2007). "Methods protocol for the human mortality database". In: *University of California, Berkeley, and Max Planck Institute for Demographic Research, Rostock. URL: https://www.mortality.org/Public/Docs/MethodsProtocol.pdf [version 31/05/2007]* 9, pp. 10–11.