MORTALITY IMPROVEMENT AND

SELECT BIRTH COHORTS

Frederik Weber

XXX Pages August 2005

This thesis provides an overview of possible criteria to measure mortality improvement and identify select cohorts. Also, results of their application to real mortality data for several countries are discussed.

APPROVED:

Date Krzysztof Ostaszewski, Chair

Date Richard MacMinn

Date Kulathavaranee Thiagarajah

MORTALITY IMPROVEMENT AND SELECT BIRTH COHORTS

Frederik Weber

XXX Pages August 2005

The twentieth century has shown unprecedented changes in mortality all over the world, causing people to live significantly longer than previously born generations. This phenomenon is commonly termed *mortality improvement* or *longevity improvement*, the former term bearing some contradiction.

However, such trends are not uniform increases, but rather cause some generations to exhibit greater improvements than others. Willets (1999) identifies such a "cohort effect" for England and Wales, describing it as "a wave of rapid improvements, rippling upwards through mortality rates". Further investigations and identification of so-called select cohorts has been been performed by Willets (2004). A more extensive review of criteria and results of their application to real data for a number of countries is given in MacMinn et al. (2005).

This thesis provides basic concepts of actuarial mathematics which are subsequently used to define a variety of possible criteria for identifying select cohorts, i.e., cohorts that exhibit significantly higher mortality improvements than other cohorts.

Furthermore, detailed results of the application of these criteria to real mortality data from the online Human Mortality Database are presented and analyzed. Common patterns for groups of countries are discussed as well as general observations inherent to specific criteria. Finally, the usefulness of the previously defined criteria is discussed with regard to the results obtained.

APPROVED:

Date Krzysztof Ostaszewski, Chair

Date Richard MacMinn

Date Kulathavaranee Thiagarajah

MORTALITY IMPROVEMENT AND SELECT BIRTH COHORTS

FREDERIK WEBER

A Thesis Submitted in Partial
Fulfillment of the Requirements
for the Degree of

MASTER OF SCIENCE

Department of Mathematics

ILLINOIS STATE UNIVERSITY

2005

THESIS APPROVED:

Date Krzysztof Ostaszewski, Chair

Date Richard MacMinn

Date Kulathavaranee Thiagarajah

ACKNOWLEDGEMENTS

The author wishes to thank the members of the thesis committee for their support throughout the creation of this thesis as well as during the joint work on investigating select birth cohorts in the related research project.

Especially, I would like to thank Krzysztof Ostaszweski for his support during the course of my studies of Actuarial Mathematics, motivating me to work on advanced topics, answering many questions and providing invaluable input.

I also wish to thank Hans-Joachim Zwiesler who has contributed to make my stay at Illinois State University possible. Besides, he provided me with many suggestions related to this thesis.

Last but not least I would like to thank my parents, Marion and Olaf Weber, for encouraging me and making it possible to study Mathematics. They also supported without hesitation my stay in the United States which helped to enhance and develop my skills and knowledge in Actuarial Sciences.

J. F. W.

CONTENTS

TABLES

FIGURES

CHAPTER I

INTRODUCTION

The twentieth century has shown drastic changes in mortality: people all over the world tend to live significantly longer than their ancestors. A slowing and eventual fade-out has often been proposed by researchers, but was equally often proven by reality to be wrong. An illustrative overview is provided by Vaupel and von Kistowski (2005).

Among actuaries, this phenomenon is commonly termed mortality improvement or longevity improvement. Both terms refer to lower mortality rates or increased life expectancies although the former term bears some contradiction.

From a personal point of view, this is a welcome development, and it also is a success partly due to public policy, successful health care systems and more conscious lifestyles. But at the same time, these mortality improvements bring new challenges to society. Extreme aging has been a trend in many countries of the northern hemisphere. For instance, between 1989 and 2002, the number of women in Germany aged 105 and over increased from 45 to 299 while the same number for males increased from 9 to 35 (Maier and Scholz, 2004).

Kinsella and Phillips (2005) provides an extended description of the aging process, differentiating in very much detail. Possible measures for aging are discussed together with a very illustrative description of specific patterns for developed and developing countries, for younger and older generations as well as societal challenges arising from this phenomenon.

It has been observed that these changes of mortality do not occur as even upward trends but rather as an uneven process. Of course, some generations benefit more than others which translates into them having unusually high life expectancies compared to less favored cohorts.

Willets (1999) performed investigations of this "cohort effect" for the United Kingdom and described it as "a wave of rapid improvements, rippling upwards through mortality rates". Cohorts experiencing outstanding mortality improvements have been named "select cohorts" and these select cohorts are not only of individual interest, but also a primary concern for pension providers.

The existence of select birth cohorts has also been described by MacMinn (2003), providing numerous years with possible cohort effects for a number of countries.

Governmental agencies running a retirement system exist in various countries as well as private insurance firms offering annuity products. In both cases, the retirement of a large number of individuals belonging to a select cohort can be quite a shock for the entire system if it is not accounted for properly: the unexpected higher longevity requires the pension provider to pay benefits over a longer period than initially assumed.

An overview of methodologies currently incorporated by life insurers to account for mortality improvement is given by Society of Actuaries (2003). However, this publication does not give an answer of what "the right" or at least possible ways are.

O'Brien (2003) gives an excellent description of implications for pension providers in the United Kingdom under a variety of scenarios related to mortality improvement. He especially notes that current regulatory requirements do not

sufficiently account for the expected trends and suggests that more prudent assumptions be incorporated.

However, the implication of the scenario of changing mortality patterns is to define possible measures for uneven changes of mortality and use them to establish criteria for identifying individual generations as select cohorts. MacMinn et al. (2005) give a more extended overview of several such criteria and results of application to real mortality data.

In Chapter II, this thesis provides a rough overview of concepts and notations used in actuarial mathematics to describe and measure mortality and longevity. This is followed by a brief section on the calculation of annuity products.

Chapter III presents the approach used by Willets (2004) to identify select cohorts. Possible changes of this criterion are illustrated by comparison of diagrams that are obtained from the application of these criteria to real data. Furthermore, possible alternative measures for mortality and longevity improvement as well as related criteria for select cohorts are defined.

Next, the previously defined criteria were applied to real mortality data for a total of eleven countries. Detailed results are shown in Chapter IV.

Chapter V then analyzes the observations made before in more detail, focusing on the select cohorts identified in each country by the different criteria. Also, common patterns are described and countries are classified into three distinct groups.

Finally, conclusions of the foregoing investigations are drawn in Chapter VI.

CHAPTER II

OVERVIEW OF ACTUARIAL MATHEMATICS

Basic Probabilities

To describe the randomness with which death occurs during an individual's life, there are several possible approaches. However, it seems natural to consider mortality as of random nature, thus treat an individual's lifetime as random as well.

In order to describe this randomness, it is convenient to introduce two closely related random variables. Of course, this concept is per se not limited to the case of a human life, but can directly be transferred to the "lifetime" of a machine or an industrial product.

Starting from these random variables, further concepts are introduced which are frequently used to describe aspects related to lifetime and mortality.

Bowers et al. (1997) provided valuable inspiration for this chapter.

Note, that unless explicitly stated, a symbol with x relates to a person of either sex. If we need to differentiate by the gender of a person, we will consider y as relating to a female person while in that case x shall relate to a male person. This differentiation is necessary since males and females exhibit very distinct patterns of mortality and gender is the main classification attribute used in the insurance industry.

Basic Random Variables

For a life-age-x, denoted by (x), let $T(x)$ be the time-until-death random variable while X is the age-at-death random variable (Bowers et al., 1997).

Note that $\mathbb{P}(...)$ instead of $\Pr(...)$ will be used to denote probabilities.

A newborn's age-at-death is a continuous random variable and its distribution function (d.f.) is

$$F_X(x) = \mathbb{P}(X \leq x) \qquad x \geq 0. \tag{2.1}$$

By

$$s_X(x) = 1 - F_X(x) = \mathbb{P}(X > x) \qquad x \geq 0 \tag{2.2}$$

we denote the survival function of X. The value $s_X(x)$ can be interpreted as the probability that a newborn will survive at least until age x (Bowers et al., 1997).

Unless necessary to distinguish the age-at-death from other random variables, we will mostly omit the subscript X when using the survival function.

Although being equivalent concepts, in actuarial mathematics the survival function is much more frequently used than the distribution function of X. Nevertheless, all formulae expressed in terms of s_X can easily be transferred to incorporate F_X using basic probability. For example, expressions such as

$$\mathbb{P}(x < X \leq z) = F_X(z) - F_X(x) = s(x) - s(z), \tag{2.3}$$

can be obtained; i.e., the probability that a newborn will die between ages x and z $(x < z)$ (Bowers et al., 1997). On the other hand,

$$\mathbb{P}(x < X \leq x + z | X > x) = \frac{F_X(x+z) - F_X(x)}{1 - F_X(x)} = \frac{s(x) - s(x+z)}{s(x)} \tag{2.4}$$

is the conditional probability that a newborn will die between ages x and $x + z$, given survival to age x (Bowers et al., 1997).

Following the International Actuarial Notation (cf. for example Bowers et al., 1997, pp.693) , we introduce for $t \geq 0$

$$ {}_tq_x = \mathbb{P}(T(x) \leq t), \tag{2.5} $$

i.e. the probability that a person aged x will die before age $x + t$, and

$$ {}_tp_x = \mathbb{P}(T(x) > t), \tag{2.6} $$

which can be regarded as the complementary probability that a person aged x will survive at least t years (Bowers et al., 1997).

Note that ${}_tp_x$ is the survival function for (x) and it is easy to to obtain the following expressions (cf. Bowers et al., 1997):

$$ T(x) = X - x \tag{2.7} $$

$$ T(0) = X \tag{2.8} $$

$$ {}_xp_0 = s(x). \tag{2.9} $$

More generalized, the conditional probability that (x) will die by age $x + t + u$, given survival to age $x + t$, is

$$ \begin{aligned} {}_t|_uq_x &= \mathbb{P}(t < T(x) \leq t + u) \\ &= {}_{t+u}q_x - {}_tq_x \\ &= {}_tp_x - {}_{t+u}p_x. \end{aligned} \tag{2.10} $$

This type of probability is also called a deferred death probability.

In the special case where $t = 1$ in formulae (2.5) and (2.6), we will omit the prefix and instead write q_x, p_x, respectively. Similarly, for $u = 1$, (2.10) will be written as ${}_t|q_x$ (Bowers et al., 1997).

At this point, it may seem that the probability of (x) to die between ages x and $x+u$ could be expressed in two ways. On the one hand, formula (2.10) with $t=0$ expresses this. But does formula (2.3) with $z=x+u$ also describe the same probability?

To answer this question, we have to account for the fact that the latter formula considers a newborn's age-at-death - while the former considers the time-until-death for a person aged x (Bowers et al., 1997). This may of course include more information than pure survival until age x.

Such additional information could be the passing of a physical examination upon the beginning of insurance or the fact that the individual's life is "special" to some extent. Whatever this information might be, it would always support the notion that the individual exhibited a different mortality than the overall population it was selected from.

For further purposes, we will not consider any additional information but simply survival to a certain age, which allows us to disregard any possible differences between formulae (2.3) and (2.10), and this permits us to write (following Bowers et al., 1997):

$$\begin{aligned} {}_{t|u}q_x &= \frac{s(x+t)-s(x+t+u)}{s(x)} \\ &= \left(\frac{s(x+t)}{s(x)}\right)\cdot\left(\frac{s(x+t)-s(x+t+u)}{s(x+t)}\right) \\ &= {}_tp_x\ {}_uq_{x+t}. \end{aligned} \tag{2.11}$$

There is another random variable, which is closely related to the future lifetime, $T(x)$. The curtate future lifetime, $K(x)$, is the number of future years

completed by (x) before death (Bowers et al., 1997). Since

$$K(x) = [T(x)], \tag{2.12}$$

$K(x)$ is the greatest integer in $T(x)$; it is a discrete random variable.

Its probability function (p.f.) is given by (Bowers et al., 1997):

$$\begin{aligned} \mathbb{P}(K(x) = k) &= \mathbb{P}(k \leq T(x) < k+1) \\ &= \mathbb{P}(k < T(x) \leq k+1) \\ &= {}_kp_x - {}_{k+1}p_x \\ &= {}_kp_x \; q_{x+k} \\ &= {}_k|q_x \qquad\qquad k = 0, 1, 2, ... \end{aligned} \tag{2.13}$$

Force of Mortality

In order to accurately describe mortality from another point of view, the notion of the force of mortality is introduced.

Formula (2.3) was established to express the conditional probability that a newborn will die between ages x and z, given survival to age x. This can, for fixed $c = z - x$, be considered as a function of x which describes the distribution of death probability in the near future, up to time c for a life that has reached the age x. We call this the instantaneous death function (see Bowers et al., 1997).

However, a more interesting result can be obtained when looking at the density of the probability of death at age x, i.e. using formula (2.3) with $z = x + \Delta x$:

$$\mathbb{P}(x < X \leq x + \Delta x) = \frac{F_X(x + \Delta x) - F_X(x)}{1 - F_X(x)} \cong \frac{f_X(x)\Delta x}{1 - F_X(x)}. \tag{2.14}$$

The last expression of (2.14) is denoted by $\mu(x)$ and can be interpreted as a conditional probability density function (Bowers et al., 1997). For each age x it gives the conditional p.d.f. of X at age x upon survival to this age. It is commonly called force of mortality or hazard rate function. Of course, the force of mortality is the negative logarithmic derivative of the survival function:

$$\mu(x) = -\frac{d}{dx} \log s(x) \tag{2.15}$$

Therefore, by applying basic integration techniques, the probability of survival can be expressed in terms of the force of mortality (Bowers et al., 1997):

$${}_np_x = \exp\left(-\int_x^{x+n} \mu(y)\ dy\right) = \exp\left(-\int_0^n \mu(x+s)\ ds\right) \tag{2.16}$$

Life Tables

The age-at-death and its related random variables are usually not described by an analytical expression, but are frequently given as tabulations by individual ages in the form of a life table.

For the purpose of establishing a life table, we follow the approach described in Bowers et al. (1997) and consider a group of l_0 newborns (a frequently used value is $l_0 = 100000$). The distribution of each individual's age-at-death is given by $s(x)$, and by $\mathcal{L}(x)$ we shall denote the number of survivors to age x within this group.

When indexing these lives by $j = 1, 2, ..., l_0$, setting

$$I_j = \begin{cases} 1 & \text{if life } j \text{ survives to age } x \\ 0 & \text{otherwise,} \end{cases} \tag{2.17}$$

and having in mind that $\mathbb{E}[I_j] = s(x)$, we can obtain

$$l_x = \mathbb{E}[\mathcal{L}(x)] = \sum_{j=1}^{l_0} \mathbb{E}[I_j] = l_0\ s(x). \tag{2.18}$$

Similarly, we have ${}_x\mathcal{D}_x$ as the number of deaths between ages x and $x+n$ from among the initial l_0 newborns, which allows us to write

$$\begin{aligned} {}_nd_x &= \mathbb{E}[{}_n\mathcal{D}_x] \\ &= l_0[s(x) - s(x+n)] \\ &= l_x - l_{x+n} \qquad (2.19) \\ d_x &= \mathbb{E}[\mathcal{D}_x] = l_x - l_{x+1}. \qquad (2.20) \end{aligned}$$

This concept of l_0 newborns with survival function $s(x)$ is called a random survivorship group (Bowers et al., 1997).

Table 1
Life table for the total population: United States, 2002

	Probability of dying between ages x to $x+1$	Number surviving to age x	Number dying between ages x to $x+1$	Person-years lived between ages x to $x+1$	Total number of person-years lived above age x	Expectation of life at age x
Age	$q(x)$	$l(x)$	$d(x)$	$L(x)$	$T(x)$	$e(x)$
0–1	0.006971	100,000	697	99,389	7,725,787	77.3
1–2	0.000472	99,303	47	99,279	7,626,399	76.8
2–3	0.000324	99,256	32	99,240	7,527,119	75.8
3–4	0.000239	99,224	24	99,212	7,427,879	74.9
4–5	0.000203	99,200	20	99,190	7,328,667	73.9
5–6	0.000176	99,180	17	99,171	7,229,477	72.9
6–7	0.000144	99,163	14	99,155	7,130,306	71.9
7–8	0.000142	99,148	14	99,141	7,031,151	70.9
8–9	0.000152	99,134	15	99,127	6,932,009	69.9
9–10	0.000145	99,119	14	99,112	6,832,883	68.9

Source: Arias (2004, p.7)

Table 1 shows an excerpt of a life table for the United States' overall population reproduced from Arias (2004, p.7). Note that other life tables may give

mortality rates within the first year of life for the first day, week, and month of life (cf. e.g. Bowers et al., 1997, pp.60).

Since there have been very few observations of human lives aged beyond 110, say, it is common to define a so-called limiting age, denoted by ω. This translates into each person having died by that age. Typical values of ω are between 100 and 110 years.

Another interpretation of a life table is the concept of a deterministic survivorship group or cohort. The basic idea is that a closed group of initially l_0 newborns is at each age x subject to annual rates of decrement, given by q_x.

From this, it follows that (see Bowers et al., 1997)

$$\begin{aligned} l_1 &= l_0(1-q_0) = l_0 - d_0 \\ l_2 &= l_1(1-q_1) = l_1 - d_1 = l_0 - (d_0 + d_1) \\ \vdots &= \vdots \\ l_x &= l_{x-1}(1-q_{x-1}) = l_{x-1} - d_{x-1} \\ &= l_0 - \sum_{y=0}^{x-1} d_y = l_0 \left(1 - \frac{\sum_{y=0}^{x-1} d_y}{l_0}\right) = l_0(1 - {}_xq_0), \end{aligned} \tag{2.21}$$

and it can be rewritten as

$$\begin{aligned} l_1 &= l_0\ p_0 \\ l_2 &= l_1\ p_1 = (l_0\ p_0)p_1 \\ \vdots &= \vdots \\ l_x &= l_{x-1}\ p_{x-1} = l_0 \left(\prod_{y=0}^{x-1} p_y\right) = l_0\ {}_xp_0 \end{aligned} \tag{2.22}$$

Other Life Table Characteristics

Apart from the basic series of values, a life table usually also tabulates some other variables which shall be introduced here. Bowers et al. (1997) provide an overview.

The complete-expectation-of-life, denoted by $\mathring{e}_x$, is the expected value of $T(x)$ and it can be expressed as

$$\begin{aligned} \mathring{e}_x &= \mathbb{E}[T(x)] \\ &= \int_0^\infty t \ {}_tp_x \ \mu(x+t) \ dt \\ &= \int_0^\infty {}_tp_x \ dt. \end{aligned} \tag{2.23}$$

Applying the same to the curtate future lifetime, we similarly get (Bowers et al., 1997):

$$\begin{aligned} e_x &= \mathbb{E}[K(x)] \\ &= \sum_{k=0}^\infty k \ {}_kp_x \ q_{x+k} \\ &= \sum_{k=1}^\infty {}_kp_x. \end{aligned} \tag{2.24}$$

The expected number of years lived between ages x and $x+1$ by survivors of the l_0 newborns is

$$L_x = \int_0^1 t \ l_{x+t} \ \mu(x+t) \ dt + l_{x+1} = \int_0^1 l_{x+t} \ dt \tag{2.25}$$

which is used in the definition of the central-death-rate (cf. Bowers et al., 1997):

$$\begin{aligned} m_x &= \frac{\int_0^1 l_{x+t} \ \mu(x+t) \ dt}{\int_0^1 l_{x+t} \ dt} \\ &= \frac{l_x - l_{x+1}}{L_x} \\ &= \frac{d_x}{L_x}. \end{aligned} \tag{2.26}$$

Both concepts can of course be applied to a n-year period as well. Thus, we obtain (Bowers et al., 1997)

$$\begin{aligned} {}_nL_x &= \int_0^n t\ l_{x+t}\ \mu(x+t)\ dt + n l_{x+n} \\ &= \int_0^n l_{x+t}\ dt \end{aligned} \tag{2.27}$$

$$\begin{aligned} {}_nm_x &= \frac{\int_0^n l_{x+t}\ \mu(x+t)\ dt}{\int_0^n l_{x+t}\ dt} \\ &= \frac{l_x - l_{x+n}}{{}_nL_x} \\ &= \frac{{}_nd_x}{{}_nL_x} \end{aligned} \tag{2.28}$$

We shall also write T_x for the total number of years lived beyond age x by the initially l_0 newborns (cf. Bowers et al., 1997), and it is defined as

$$T_x = \int_0^\infty t\ l_{x+t}\ \mu(x+t)\ dt = \int_0^\infty l_{x+t}\ dt, \tag{2.29}$$

and it can be related to the complete life expectancy (2.23) by

$$\mathring{e}_x = \frac{\int_0^\infty l_{x+t}\ dt}{l_x} = \int_0^\infty {}_tp_x\ dt = \frac{T_x}{l_x}. \tag{2.30}$$

Assumptions for Fractional Ages

One further aspect of mortality rates needs to be considered. Although some life tables may give values for the first months of life (or even shorter periods, as can be seen in Table 1), the usual time period is a year. However, we sometimes need to specifically account for the fact that death or survival do not exactly match whole multiples of a year.

For fractional ages, the basic idea is to interpolate the values of the survival function $s(x)$ at the neighboring whole numbers of years. The most frequent

assumptions are linear, exponential, and harmonic interpolation, although there is variety of other ways to obtain values for fractional ages (see e.g. Bowers et al., 1997).

Due to its simplicity and frequent application, as well as in consideration of the fact that we will later on focus on it, we will focus on the linear interpolation:

$$s(x+t) = (1-t)s(x) + ts(x+1) \qquad 0 \leq t \leq 1, \tag{2.31}$$

which is also known as the uniform distribution of deaths (UDD) assumption within each year of age (Bowers et al., 1997).

The UDD implies that ${}_tp_x$ is a linear function of t (Bowers et al., 1997), and we can also obtain

$$\begin{aligned}
{}_tq_x &\overset{UDD}{=} t\ q_x && (2.32)\\
\mu(x+t) &\overset{UDD}{=} \frac{q_x}{1-t\ q_x} && (2.33)\\
{}_tp_x &\overset{UDD}{=} 1-t\ q_x && (2.34)\\
{}_tp_x\ \cdot\ \mu(x+t) &\overset{UDD}{=} q_x. && (2.35)
\end{aligned}$$

Analytical Laws of Mortality

Although statisticians may be much more familiar with an analytical distribution or survival function of a random variable, this is - in general - not the case in actuarial mathematics, where the respective values are usually taken from tabulations, such as a life table (Bowers et al., 1997).

Yet, it sometimes is more convenient to assume that the randomness of mortality can be described analytically.

There are several families of functions (cf. Bowers et al., 1997), also called

laws of mortality, each bearing practical advantages and disadvantages.

The oldest such law of mortality, known as De Moivre's law, was published as early as 1729, proposing (Bowers et al., 1997)

$$\begin{aligned} \mu(x) &= \frac{1}{\omega - x} \\ s(x) &= 1 - \frac{x}{\omega} \qquad 0 \leq x < \omega. \end{aligned} \tag{2.36}$$

In 1825, Gompertz proposed a law of mortality of the form

$$\begin{aligned} \mu(x) &= Bc^x \\ s(x) &= \exp(-\frac{B}{\log c}(c^x - 1)) \qquad B > 0, c > 1, x \geq 0 \end{aligned} \tag{2.37}$$

which was generalized by Makeham in 1860 (Bowers et al., 1997) to

$$\begin{aligned} \mu(x) &= A + Bc^x \\ s(x) &= \exp(-Ax - \frac{B}{\log c}(c^x - 1)) \qquad B > 0, A \geq -B, c > 1, x \geq 0. \end{aligned} \tag{2.38}$$

Especially the Gompertz law has been widely used in actuarial sciences, whenever an analytical law of mortality was used. However, Yue (2002) mentions that there are some difficulties when approximating mortality rates for the oldest-old population. Having in mind, that especially elderly mortality improvements are significant (cf. Willets, 1999), the Gompertz law should be used with caution when modeling real data. Possibly, adaptations or separate assumptions for older ages have to be made.

Weibull (1939) assumed a different class of functions (Bowers et al., 1997):

$$\begin{aligned} \mu(x) &= kx^n \\ s(x) &= \exp(-\frac{k}{n+1}x^{n+1}) \qquad k > 0, n > 0, x \geq 0. \end{aligned} \tag{2.39}$$

However, for the purpose of this thesis, these analytical laws shall be neglected in favor of the consideration of life tables.

Life Annuities

Apart from the basic probabilities and concepts presented in the previous section, we will also be looking at life annuities in this thesis. Hence we give a brief introduction to the actuarial concepts behind them, following Bowers et al. (1997).

Life annuities are a series of payments contingent on survival – as opposed to annuities-certain which are widely used in interest theory whose payments are theoretically unconditional (Bowers et al., 1997).

The annuity payments may be temporary or payable for the whole life, and they also may be deferred by a number of years. Furthermore, payments can be due at the beginning of each interval considered, or at the respective end of each period. In the former case they are called annuities-due, while in the latter case we are dealing with annuities-immediate (Bowers et al., 1997). Note that we will mostly assume intervals of one year; but this is not restrictive.

Annuities are used in the calculation of life insurance products where premiums are usually paid in the form of an annuity-due. But their most important role is in the actuarial calculation of pension products, which do not only provide a lump sum, but regular payments.

A whole life annuity-due pays a certain amount, mostly one monetary unit (hence often called unit annuity), at the beginning of each year as long as the individual is alive. Its present value is a random variable (see Bowers et al., 1997), denoted by

$$Y = \ddot{a}_{\overline{K+1}|} \tag{2.40}$$

where K is the curtate future lifetime of x, which was defined as a discrete random variable in (2.12), and its probability function is given by (2.13). Note that we use $K + 1$ instead of K, because the payments are made at the beginning of each year.

The actuarial present value, or expected present value of future payments of such an annuity is (Bowers et al., 1997)

$$\begin{aligned}
\ddot{a}_x &= \mathbb{E}[Y] = \mathbb{E}[\ddot{a}_{\overline{K+1}|}] \\
&= \sum_{k=0}^{\infty} \ddot{a}_{\overline{k+1}|}\ {}_kp_x\ q_{x+k} \\
&= 1 + \sum_{k=0}^{\infty} v^{k+1}\ {}_{k+1}p_x \\
&= \sum_{k=0}^{\infty} v^k\ {}_kp_x, \qquad (2.41)
\end{aligned}$$

and by applying basic relations from the beginning of this chapter we can obtain

$$\begin{aligned}
\ddot{a}_x &= 1 + \sum_{k=0}^{\infty} v^{k+1}\ {}_{k+1}p_x \\
&= 1 + v\ p_x \sum_{k=0}^{\infty} v^k\ {}_kp_{x+1} \\
&= 1 + v\ p_x \ddot{a}_{x+1}. \qquad (2.42)
\end{aligned}$$

The same idea can of course be applied to the case where the annuity payments end after n years; this is called a n-year temporary life annuity-due (Bowers et al., 1997). The corresponding present value random variable is

$$Y = \begin{cases} \ddot{a}_{\overline{K+1}|} & 0 \le K < n \\ \ddot{a}_{\overline{n}|} & K \ge n \end{cases} \qquad (2.43)$$

while its actuarial present value is given by

$$\begin{aligned}
\ddot{a}_{x:\overline{n}|} &= \mathbb{E}[Y] \\
&= \sum_{k=0}^{n-1} \ddot{a}_{\overline{k+1}|}\ {}_kp_x\ q_{x+k} + \ddot{a}_{\overline{n}|}\ {}_np_x \\
&= \sum_{k=0}^{n-1} v^k\ {}_kp_x. \qquad (2.44)
\end{aligned}$$

The corresponding recursion formula (Bowers et al., 1997) is

$$\ddot{a}_{x:\overline{y-x}|} = 1 + v\ p_x\ \ddot{a}_{x+1:\overline{y-(x+1)}|} \tag{2.45}$$

Since some considerations in later chapters will be based on the actuarial present value of an annuity-immediate, we relate it to an annuity-due. An annuity-immediate pays at the end of each year, given survival of the individual to the beginning of the respective year (see Bowers et al., 1997). Hence we have

$$\begin{aligned} a_x &= \mathbb{E}[a_{\overline{K}|}] \\ &= \sum_{k=0}^{\infty} {}_kp_x\ q_{x+k}\ a_{\overline{k}|} \\ &= \sum_{k=1}^{\infty} v^k\ {}_kp_x. \end{aligned} \tag{2.46}$$

Similarly, for an n-year temporary annuity-immediate we have (Bowers et al., 1997)

$$Y = \begin{cases} a_{\overline{K}|} & 0 \le K < n \\ a_{\overline{n}|} & K \ge n \end{cases} \tag{2.47}$$

$$\begin{aligned} a_{x:\overline{n}|} &= \mathbb{E}[Y] \\ &= \sum_{k=1}^{n} v^k\ {}_kp_x. \end{aligned} \tag{2.48}$$

The two types of annuities are related by (Bowers et al., 1997)

$$\ddot{a}_x = 1 + a_x \tag{2.49}$$

$$\ddot{a}_{x:\overline{n}|} = 1 + a_{x:\overline{n-1}|} \tag{2.50}$$

This concludes the brief introduction to the calculation of life annuities as far as their application in this thesis is concerned. The calculation of other types of annuities is treated in greater detail e.g. in Bowers et al. (1997).

CHAPTER III

DEFINING AND IDENTIFYING SELECT COHORTS

Introductory Questions and Discussion

Before we consider and compare definitions for select cohorts, let us have a look at some basic questions arising when talking about "select cohorts".

The first question might be what a select cohort is. The word "cohort" has already been mentioned in the context of the deterministic survivorship group approach (cf. p.11). It basically describes a "group of individuals with a common birth period" (Bowers et al., 1997, p. 586). A more detailed definition is found in Wikipedia (2005), describing it as "a group of subjects — most often humans from a given population — defined by a condition on their date of birth".

We will almost always interpret the birth period as a calendar year which means we will consider generations born in different calendar years. The main reason is that most demographic and mortality data is available by calendar year.

The other term that needs to be defined here is the word "select". In this context, it describes a cohort as special, somewhat outstanding compared to other cohorts - with respect to mortality or longevity. This always includes unusual high longevity, or significantly lower mortality rates, but we will later on also present other specifications for a cohort being "select". Of course, precise definitions of a select cohort still need to be given.

Another question that might arise at this point is why we would want to establish a definition (or even multiple definitions) for select cohorts. The answer is

clearly the same as in many other fields of research: a cohort should not be arbitrarily declared "select", but instead we want to have a set of mathematically precise criteria which allow us to exactly determine select cohorts when investigating (demographic) data.

The purpose of establishing more than one criterion is to enable us to investigate possible differences among these criteria as well as discuss strengths and weaknesses of definitions. As will turn out later, different definitions need not coincide; they can in fact produce different results.

Furthermore, we could also briefly think about some naive ideas of how to determine select cohorts. For instance, life expectancies at birth could be compared for several cohorts. But this would neglect possible changes of life expectancy during life time. Besides, it does not respond to the special concerns of public and private pension providers. These entities will very likely be concerned about an increase of a cohort's longevity during their retirement period.

Changes of mortality rates could be another approach - either at a specific age, during a period in life, or as some type of aggregation across ages. Again, this criterion would not be very precise, not accounting for subtle changes during the course of a life.

One last concern should be raised at this point before actually discussing definitions: Even if it is clear which demographic indices shall be considered, it is not obvious which cohorts are to be compared. For instance, a certain cohort could have (in any sense) "better" values than some other cohort born some decades ago. But at the same time, this distinction may not hold for this cohort when being compared to an adjacent cohort. Also, the definition of a "global" criterion has to be handled with care, since this clearly depends on the range of available data, and

might change whence more data can be obtained.

Regardless of which approach is chosen, the price to be paid for obtaining a single index or ratio (or a small number thereof) is that some of the details are lost. Hence, advantages and disadvantages have to be carefully taken into consideration.

Existing Definitions and Modifications

A Definition Proposed by Willets

Recall the Gompertz law of mortality (2.37), p.15, which says that $\mu(x)$ is a log-linear function of the age x, i.e. $\ln(\mu(x))$ is linear in x – or more formally: $\ln(\mu(x)) = a + b{\cdot}x$.

To estimate the force of mortality, $\mu(x)$, we can use the central death rate m_x (see Bowers et al., 1997). This is an interesting and useful fact, because the central death rate is a frequently used concept, especially in demography - as opposed to the mortality rate q_x. This relation between the force of mortality and the central death rate enables us to consider $\ln(m_x)$ as a linear function of age as well.

Willets (2004) uses this approach together with a smoothing technique to obtain rates of mortality improvement. Compare the following description also to MacMinn et al. (2005).

To distinguish between central death rates for different cohorts, let ${}^{z}m_x$ denote the central death rate for (x), born in calendar year z. This implies that we are considering data in the specific calendar year $z + x$.

First, a log-linear regression of central death rates over a period of $\pm$ 4 years was performed. For instance, the estimation of ${}^{z}m_x$ for calendar year $z + x$ is obtained from running a regression on the respective values for calendar years

$(z+x-4, z+x-3, z+x-2, z+x-1, z+x, z+x+1, z+x+2, z+x+3, z+x+4)$. Obviously, this is the same as considering cohorts born in
$(z-4, z-3, z-2, z-1, z, z+1, z+2, z+3, z+4)$.

Since $\ln(m_x)$ is considered a linear function of x, the regression gives us the relationship

$$\ln({}^{z}m_x) = \alpha(x,z) + \beta(x,z) \cdot (z+x) + \varepsilon(x,z), \tag{3.1}$$

where $\varepsilon(x,z)$ is the residual term with $\varepsilon(x,z) \sim N(0,1)$ and $\alpha(x,z)$ and $\beta(x,z)$ are the regression parameters (MacMinn et al., 2005). Estimating the parameters yields

$$\ln({}^{z}\hat{m}_x) = \hat{\alpha}(x,z) + \hat{\beta}(x,z) \cdot (z+x). \tag{3.2}$$

The value for the preceding cohort is obtained in a similar fashion, resulting in

$$\begin{aligned} {}^{z-1}\hat{m}_x &= e^{\hat{\alpha}(x,z-1)+\hat{\beta}(x,z-1)\cdot(z+x-1)} && (3.3)\\ {}^{z}\hat{m}_x &= e^{\hat{\alpha}(x,z)+\hat{\beta}(x,z)\cdot(z+x)} && (3.4) \end{aligned}$$

To actually measure mortality improvement, we consider differences of the central death rates, as obtained through the above estimations. Since lower mortality should produce a positive value, and given the fact that absolute numbers are hard to compare, we define the relative rate of mortality improvement (MacMinn et al., 2005) as

$$\begin{aligned} {}^{z}\Delta\tilde{m}_x &= \frac{{}^{z-1}\hat{m}_x - {}^{z}\hat{m}_x}{{}^{z-1}\hat{m}_x} \\ &= 1 - \frac{{}^{z}\hat{m}_x}{{}^{z-1}\hat{m}_x} \\ &= 1 - \frac{e^{\hat{\alpha}(x,z)+\hat{\beta}(x,z)\cdot(z+x)}}{e^{\hat{\alpha}(x,z-1)+\hat{\beta}(x,z-1)\cdot(z+x-1)}} \end{aligned} \tag{3.5}$$

At this point, Willets' calculations are siginficantly simplified. He takes the same estimations for α and β for the previous cohort instead of re-estimating them

based on the data for the calendar years

$(z+x-5, z+x-4, z+x-3, z+x-2, z+x-1, z+x, z+x+1, z+x+2, z+x+3)$

which yields a simplified version of (3.5) (see also MacMinn et al., 2005):

$$ {}^{z}\Delta\tilde{m}_x = 1 - \frac{e^{\hat{\alpha}(x,z)+\hat{\beta}(x,z)\cdot(z+x)}}{e^{\hat{\alpha}(x,z)+\hat{\beta}(x,z)\cdot(z+x-1)}} = 1 - e^{\hat{\beta}(x,z)} \tag{3.6} $$

Then, Willets compares these values in a given calendar year among all the cohorts alive in that year - each at a different age. For instance, for calendar year z, he compares the mortality improvement rates $(..., {}^{z-1}\Delta\tilde{m}_{x+1}, {}^{z}\Delta\tilde{m}_x, {}^{z+1}\Delta\tilde{m}_{x-1}, ...)$. Note that prefix and postfix always add up to $z + x$, the calendar year considered.

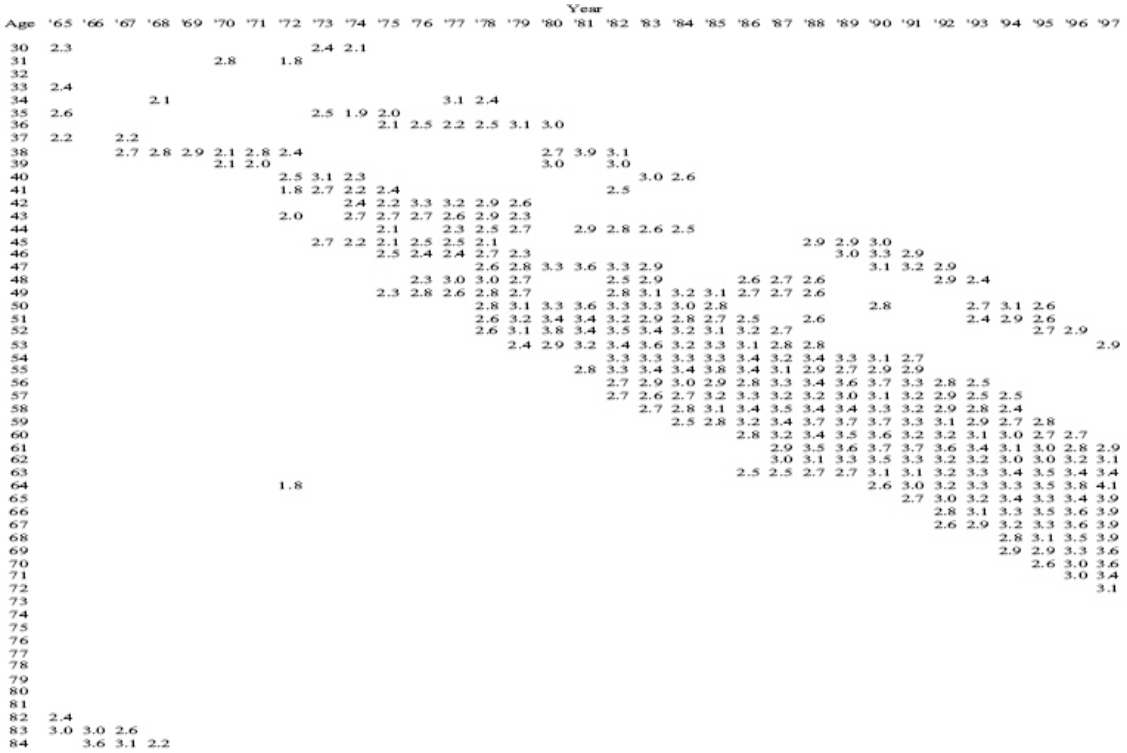
Values that exceed a certain percentage of the maximum value of such a series (for calendar year $z + x$) are considered to be "special" and hence, might support the notion of that specific cohort being select.

Willets considers those cohorts select cohorts whose respective mortality improvement rates still appear for a consecutive number of ages. This can visually be described as a diagonal pattern of a "northwest-southeast direction".

Figure 1, p. 24, shows such a diagram. For each column (calendar years between 1965 and 1997) and row (age between 30 years and 85 years), the percentage mortality improvement rate for the male population in England and Wales is shown, if that number exceeds the threshold of 70% of the per-column maximum. The underlying data was obtained through the smoothing process described above over a period of nine years, i.e. $\pm$ 4 years. In this diagram such a diagonal accumulation of printed, i.e. outstanding, mortality improvement rates can very well be identified.

While this graphical representation and identification of select cohorts is not a precise algorithm (in a mathematical sense), it is quite instructive to raise and

Figure 1
Average rates of mortality improvement



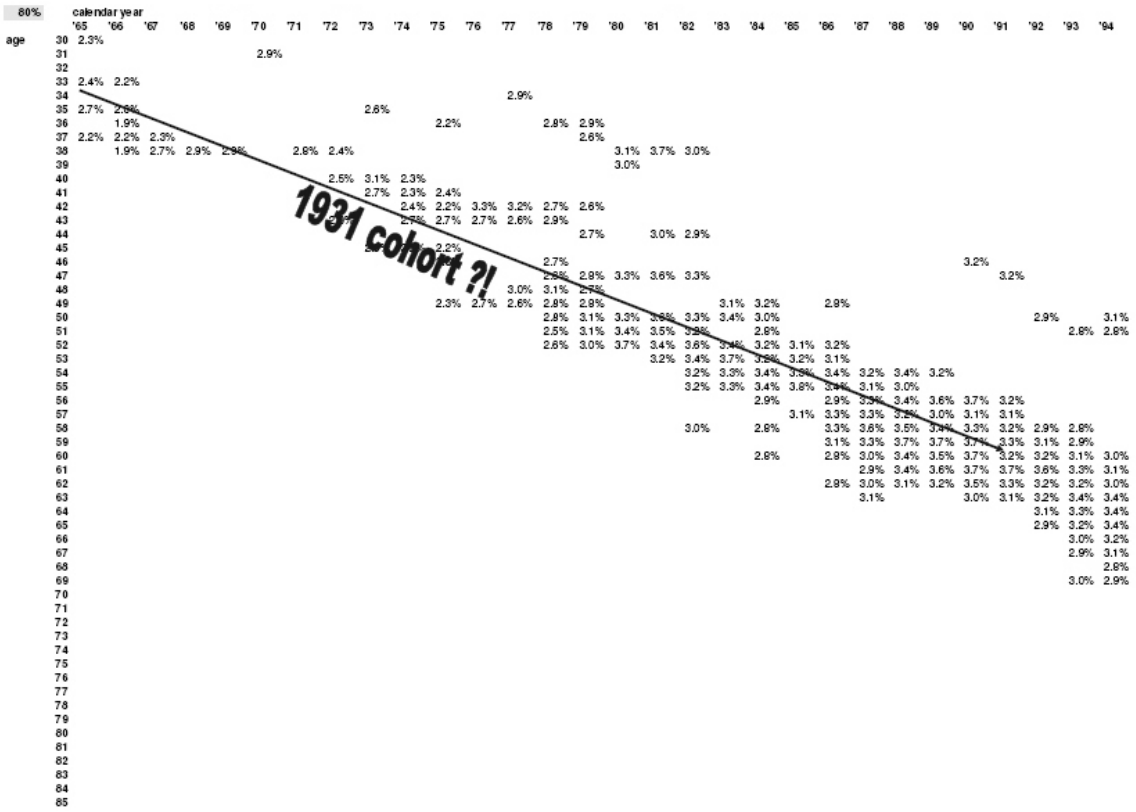
Source: Willets (2004, p. 42)

lower the threshold percentage to identify those cohorts whose values are still shown for even higher percentages, which is equivalent to determining the strongest select cohorts.

By increasing the threshold percentage the number of values that appear in the tabular scheme decreases. Figure 2, p. 25, shows the same data as Figure 1, but uses a threshold of 80% - instead of 70% in Willets' original diagram. In addition, a line shows those values that would correspond to the cohort born in 1931. Willets identified this cohort as select.

Figure 2

Average rates of mortality improvement (threshold increased to 80%)

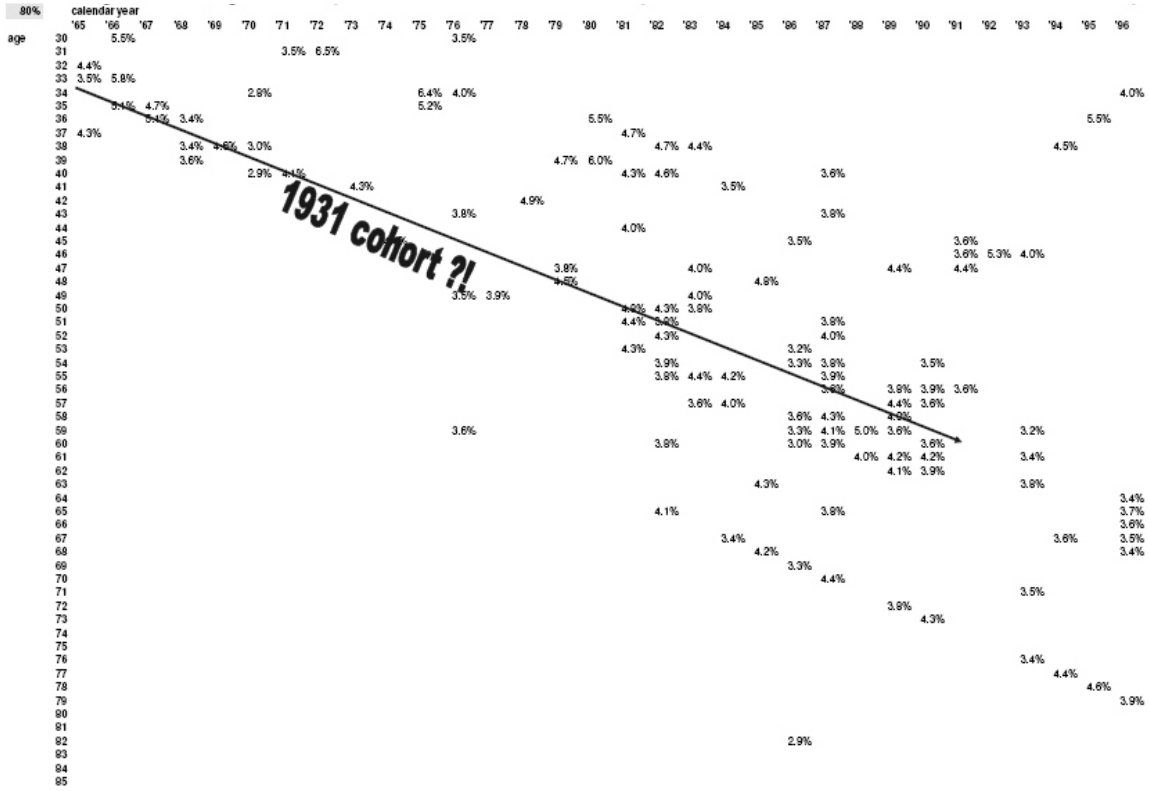


Data Source: Human Mortality Database (2005); own calculation

Also, by narrowing the smoothing period from nine years, i.e. ± 4 years, to

five or three years, changes in the patterns can be identified. Figures 3, p. 26, and 4, p. 27, show once again the same data, but with decreased smoothing periods of five and three years, respectively.

Figure 3
Average rates of mortality improvement
(smoothing period decreased to five years)

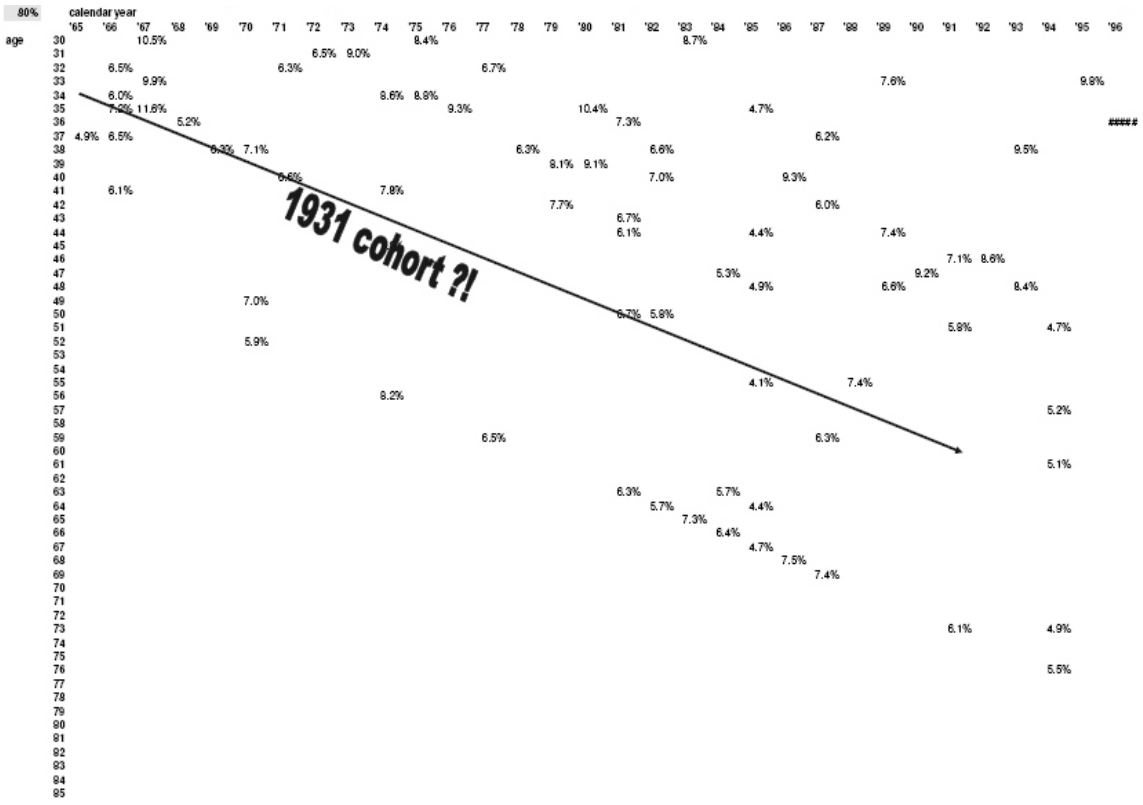


Data Source: Human Mortality Database (2005); own calculation

Even if it is hard to conclude anything at all in this case, it is almost obvious that the diagonal pattern which could be observed in Figures 1 and 2 seem to disappear whence the smoothing period is narrowed.

Apart from being a vague definition, the idea behind this criterion itself may be criticized (MacMinn et al., 2005). By defining a threshold per column, i.e. over

Figure 4
Average rates of mortality improvement
(smoothing period decreased to three years)



Data Source: Human Mortality Database (2005); own calculation

some calendar year, Willets compares mortality improvement rates of different cohorts at different ages.

As it seems more natural to compare one cohort to another at the same respective ages, the threshold applied by Willets can be replaced by a certain percentage of the maximum value (or percentile) over all cohorts at a specific age. This corresponds to defining a threshold across each row and comparing each value of the row to it. This idea is discussed more detailed in MacMinn et al. (2005).

Not only the definition has changed, but also the generated patterns under this alternative criterion. Some cohorts, previously clearly identified as select cohorts, now bear no special behavior when compared to other cohorts.

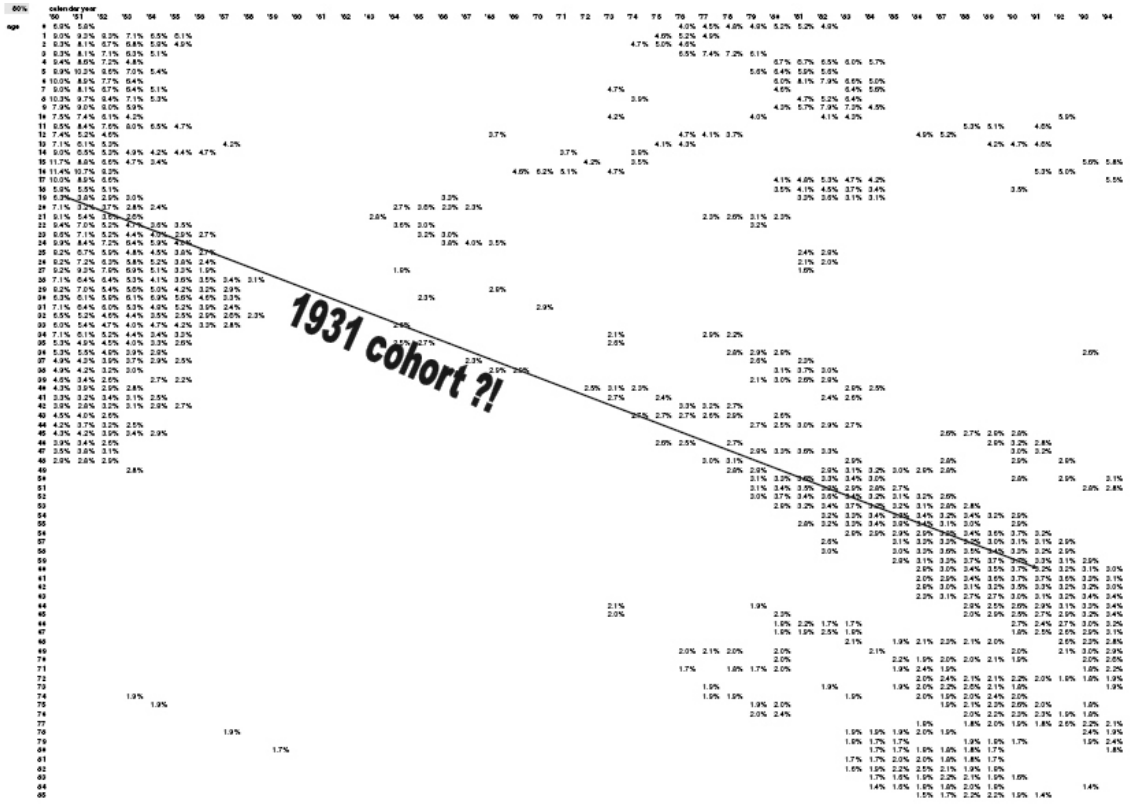
Figures 5, 6, and 7, pp. 27, show the same data as before with smoothing periods of nine, five, and three years, respectively - but with the threshold for showing or not showing a value being replaced by a by-row percentile.

With this alternative threshold, the observation is the opposite of what it was before: the shorter the smoothing period the clearer distinct select cohorts appear, i.e. diagonal patterns become visible.

However, the specific select cohort born in 1931 that was discussed by Willets (2004) does not seem to be specifically outstanding. This rather seems to be the case with cohorts born in later years.

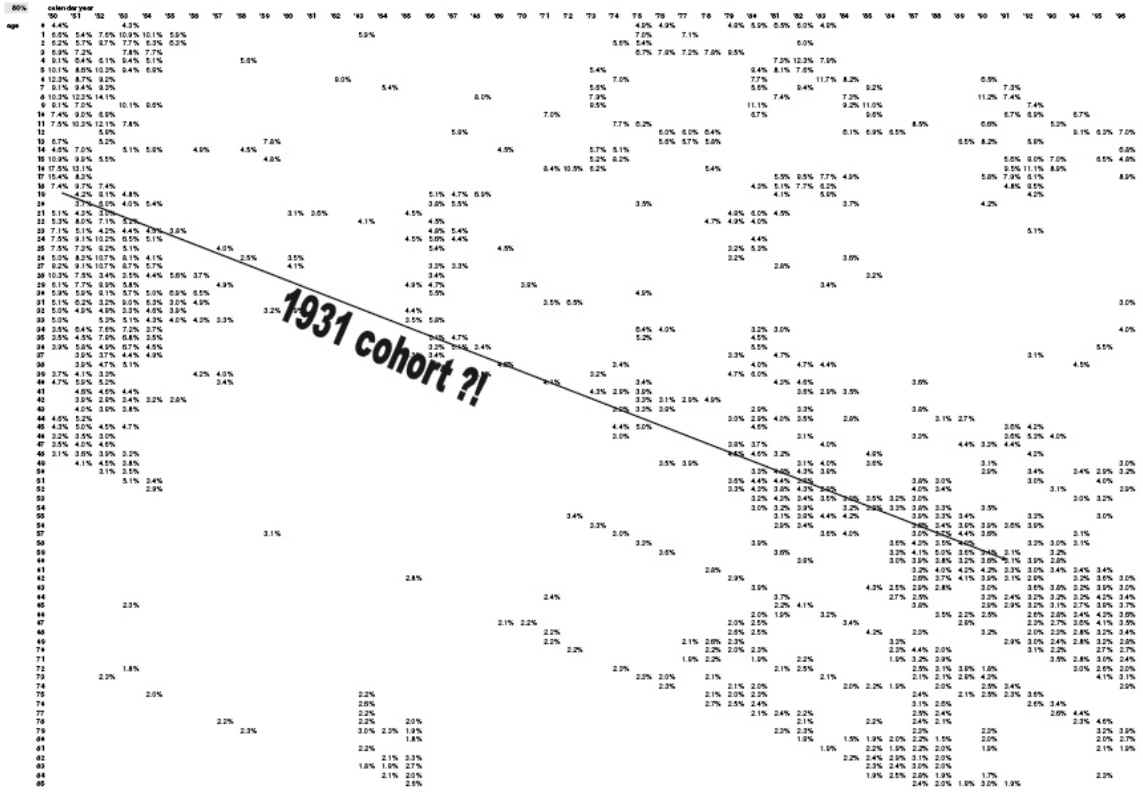
While it may be confusing that the two approaches described above give completely different result, it is important to understand the different ideas behind them. The first approach, as suggested by Willets (2004), aims at the impact that longevity or mortality improvement could have in a certain calendar year while the alternative criterion stresses the impact of a certain generation.

Figure 5
Average rates of mortality improvement
(alternative threshold, nine years of smoothing period)



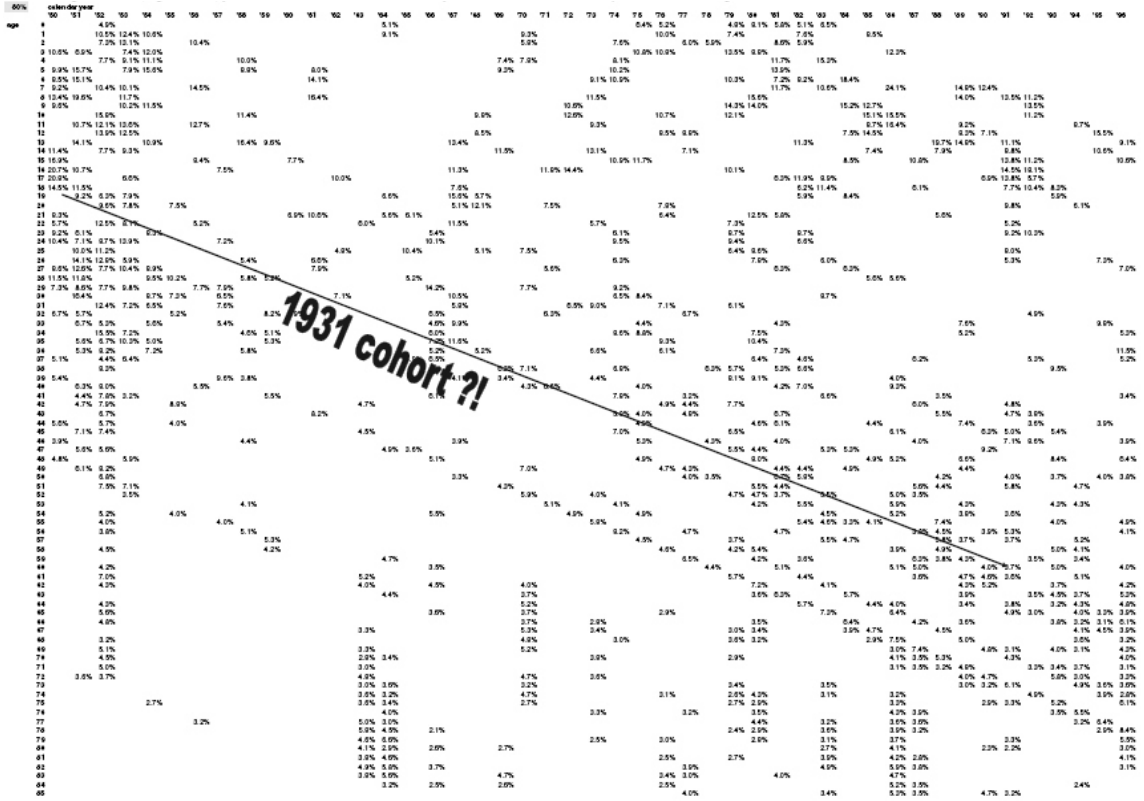
Data Source: Human Mortality Database (2005); own calculation

Figure 6
Average rates of mortality improvement
(alternative threshold, five years of smoothing period)



Data Source: Human Mortality Database (2005); own calculation

Figure 7
Average rates of mortality improvement
(alternative threshold, three years of smoothing period)



Data Source: Human Mortality Database (2005); own calculation

The former approach can be considered an answer to questions typically raised by governmental agencies in charge of a retirement insurance or compensation. Such an agency may be much more interested to know in which calendar year extraordinary high mortality improvement rates are to occur among the retirees than which generation specifically exhibits these improvement rates.

On the other hand, the latter approach is related to the situation of a private insurance company engaged in life annuities. Since pricing of a life annuity is done on an individual basis, expected higher mortality improvement rates of a specific cohort could be reflected in higher premiums for members of such a cohort.

A possible modification of Willets' approach consists in comparing mortality improvement rates not in a certain calendar year $z + x$, but rather at a given age x: $(\dots, {}^{z-1}\Delta\tilde{m}_x, {}^{z}\Delta\tilde{m}_x, {}^{z+1}\Delta\tilde{m}_x, \dots)$. This approach would also compare different cohorts, but unlike Willets not at different ages. We will refer to this idea later on.

Alternative Approaches

Since the approach proposed by Willets does not give an exact definition, but rather a hint whether a specific cohort could be a select cohort, we shall now focus on criteria which directly compare cohorts.

Select Cohorts with Respect to Mortality Improvement

Following Willets' definition, but without performing a smoothing of the underlying mortality data, we define "unsmoothed" mortality improvement rates (MacMinn et al., 2005).

Definition 1 (Mortality improvement rate)

$$ {}^{z-x}im_x = \frac{{}^{z-x-1}m_x - {}^{z-x}m_x}{{}^{z-x-1}m_x} = 1 - \frac{{}^{z-x}m_x}{{}^{z-x-1}m_x} \tag{3.7} $$

$\diamond$

If for a specific cohort these rates are frequently enough (i.e. for a sufficiently large number of ages) higher than the corresponding values of the two adjacent cohorts, we will consider this specific cohort "select". More formally (MacMinn et al., 2005):

Definition 2 (Select cohort with respect to mortality improvement)
If, for a certain cohort born in year z, at least half of mortality improvement rates are higher than the corresponding values of the two adjacent cohorts born in years $z-1$ and $z+1$, respectively, i.e.,

$$ \frac{\#\{x | {}^{z-1}im_x \leq {}^{z}im_x \geq {}^{z+1}im_x\}}{\#\{available \;\; {}^{z}im_x\}} \;\geq \frac{1}{2}, \tag{3.8} $$

we say that the cohort born in z is a ***select cohort with respect to mortality improvement****. (Note, that $\#\{...\}$ denotes the number of elements.)*
If the ratio in (3.8) *is at least $p\%$, we say that it is a* ***select cohort at*** *$p\%$.* $\diamond$

The number of available mortality improvement rates is of course restricted by the underlying data set. It may very well be that it is possible to investigate data only for a restricted range of ages, say 63-110 years.

Naturally, this gives rise to the question whether different select cohorts can be compared at all if the ranges of ages differ. Ideally, there is sufficient data

available to base a decision of a cohort being select or not on a large number of ages. At the same time, this is clearly not the case for more recent generations because only when all of the individuals have died mortality data becomes complete.

Still, there might be most interest in the effects of mortality improvement and the possible existence of select cohorts for recently born generations whose members have not yet completely died. At this point, a certain degree of incompleteness of the data simply has to be accepted.

To visualize the application of the previously described criterion, consider an example taken from MacMinn et al. (2005).

Example 1 (Population of guinea pigs (MacMinn et al., 2005))
Consider a hypothetical population of five generations of five guinea pigs, born in years 1999, 1998, 1997, 1996 and 1995, with the following central death rates at ages 1 through 5 (we assume that $\omega = 5$ is the limiting age for these guinea pigs):

$$\begin{bmatrix} {}^{1995}m_0 & {}^{1996}m_0 & {}^{1997}m_0 & {}^{1998}m_0 & {}^{1999}m_0 \\ {}^{1995}m_1 & {}^{1996}m_1 & {}^{1997}m_1 & {}^{1998}m_1 & {}^{1999}m_1 \\ {}^{1995}m_2 & {}^{1996}m_2 & {}^{1997}m_2 & {}^{1998}m_2 & {}^{1999}m_2 \\ {}^{1995}m_3 & {}^{1996}m_3 & {}^{1997}m_3 & {}^{1998}m_3 & {}^{1999}m_3 \\ {}^{1995}m_4 & {}^{1996}m_4 & {}^{1997}m_4 & {}^{1998}m_4 & {}^{1999}m_4 \end{bmatrix} = \begin{bmatrix} \frac{4}{17} & \frac{4}{17} & \frac{4}{19} & \frac{4}{20} & \frac{4}{20} \\ \frac{4}{13} & \frac{4}{13} & \frac{4}{14} & \frac{4}{15} & \frac{4}{15} \\ \frac{4}{9} & \frac{4}{9} & \frac{4}{11} & \frac{4}{11} & \frac{4}{12} \\ \frac{4}{5} & \frac{4}{5} & \frac{4}{7} & \frac{4}{8} & \frac{4}{8} \\ 2 & 2 & \frac{4}{3} & 1 & \frac{4}{3} \end{bmatrix} \quad (3.9)$$

This set of data corresponds to the following ages of death:

Birth year 1995	0.25	1.25	2.25	3.25	4.50
Birth year 1996	0.25	1.25	2.25	3.25	4.50
Birth year 1997	0.75	1.50	2.75	3.75	4.75
Birth year 1998	1.00	1.75	2.75	4.00	5.00
Birth year 1999	1.00	1.75	3.00	4.00	4.75

Using mortality rates, all five cohorts follow the simple De Moivre's Law (2.36) *with limiting age of 5, and the mortality rates do not improve among them at*

all. By using central death rates, we obtain the following mortality improvement rates for the 1996, 1997, 1998, and 1999 cohorts:

$$\begin{bmatrix} {}^{1996}im_0 & {}^{1997}im_0 & {}^{1998}im_0 & {}^{1999}im_0 \\ {}^{1996}im_1 & {}^{1997}im_1 & {}^{1998}im_1 & {}^{1999}im_1 \\ {}^{1996}im_2 & {}^{1997}im_2 & {}^{1998}im_2 & {}^{1999}im_2 \\ {}^{1996}im_3 & {}^{1997}im_3 & {}^{1998}im_3 & {}^{1999}im_3 \\ {}^{1996}im_4 & {}^{1997}im_4 & {}^{1998}im_4 & {}^{1999}im_4 \end{bmatrix} = \left[\begin{array}{c||c||cc} 0 & \frac{2}{19} & \frac{1}{20} & 0 \\ 0 & \frac{1}{14} & \frac{1}{15} & 0 \\ 0 & \frac{2}{11} & 0 & \frac{1}{12} \\ 0 & \frac{2}{7} & \frac{1}{8} & 0 \\ 0 & \frac{1}{3} & \frac{1}{4} & -\frac{1}{3} \end{array}\right] \tag{3.10}$$

An observation of the mortality improvement rates shows that the cohort born in 1997 clearly is a select birth cohort with respect to mortality improvement (at 50%). It is highlighted in (3.10)*.* ◇

However, this example is hypothetical and an extreme case where - at all ages - a cohort has higher mortality improvement rates than the adjacent cohorts. It therefore seems to be a good idea to provide an alternative definition to account for less pronounced situations of a cohort being "select".

Select Cohorts with Respect to Longevity Improvement

The concept of considering mortality improvement rates at each age and comparing them to the values of adjacent cohorts can also be applied to another important index describing mortality: life expectancy.

Recall the definition of the complete life expectancy $\mathring{e}_x$ in (2.23), p. 12, which is the key index used to describe longevity. Proceeding as before, we can define longevity improvement. To distinguish between cohorts, we shall also include a prefix denoting the year of birth (see MacMinn et al., 2005).

Definition 3 (Longevity improvement rate)

For a cohort born in calendar year z, we define the ***longevity improvement rate at age*** *x as*

$$ {}^{z}ie_x = \frac{{}^{z}\mathring{e}_x - {}^{z-1}\mathring{e}_x}{{}^{z-1}\mathring{e}_x}. \tag{3.11} $$

$\diamond$

By taking the difference in this order (as opposed to the definition of mortality improvement rates), we ensure that positive values actually describe the "positive" situation of higher life expectancy.

Following Definition 2, we apply the same idea to longevity improvement rates:

Definition 4 (Select cohort with respect to longevity improvement)

If, for a certain cohort born in year z, at least half of longevity improvement rates are higher than the corresponding values of the two adjacent cohorts born in years $z-1$ and $z+1$, respectively, i.e.,

$$ \frac{\#\{x | {}^{z-1}ie_x \leq {}^{z}ie_x \geq {}^{z+1}ie_x\}}{\#\{available \;\; {}^{z}ie_x\}} \;\geq \frac{1}{2}, \tag{3.12} $$

we say that the cohort born in z is a ***select cohort with respect to longevity improvement****.*

If the ratio in (3.12) *is at least $p\%$, we say that it is a* ***select cohort at*** *$p\%$.* $\diamond$

To illustrate the application of this criterion to actual data we use again the hypothetical population of guinea pigs from Example 1.

Example 2 (Population of guinea pigs (MacMinn et al., 2005), continued)
Consider again the population of guinea pigs. Their complete life expectancies are given by

$$\begin{bmatrix} {}^{1995}\mathring{e}_0 & {}^{1996}\mathring{e}_0 & {}^{1997}\mathring{e}_0 & {}^{1998}\mathring{e}_0 & {}^{1999}\mathring{e}_0 \\ {}^{1995}\mathring{e}_1 & {}^{1996}\mathring{e}_1 & {}^{1997}\mathring{e}_1 & {}^{1998}\mathring{e}_1 & {}^{1999}\mathring{e}_1 \\ {}^{1995}\mathring{e}_2 & {}^{1996}\mathring{e}_2 & {}^{1997}\mathring{e}_2 & {}^{1998}\mathring{e}_2 & {}^{1999}\mathring{e}_2 \\ {}^{1995}\mathring{e}_3 & {}^{1996}\mathring{e}_3 & {}^{1997}\mathring{e}_3 & {}^{1998}\mathring{e}_3 & {}^{1999}\mathring{e}_3 \\ {}^{1995}\mathring{e}_4 & {}^{1996}\mathring{e}_4 & {}^{1997}\mathring{e}_4 & {}^{1998}\mathring{e}_4 & {}^{1999}\mathring{e}_4 \end{bmatrix} = \begin{bmatrix} \frac{23}{10} & \frac{23}{10} & \frac{27}{10} & \frac{29}{10} & \frac{29}{10} \\ \frac{29}{16} & \frac{29}{16} & \frac{35}{16} & \frac{19}{8} & \frac{19}{8} \\ \frac{4}{3} & \frac{4}{3} & \frac{7}{4} & \frac{23}{12} & \frac{23}{12} \\ \frac{7}{8} & \frac{7}{8} & \frac{5}{4} & \frac{3}{2} & \frac{11}{8} \\ \frac{1}{2} & \frac{1}{2} & \frac{3}{4} & 1 & \frac{3}{4} \end{bmatrix} \tag{3.13}$$

and by applying Definition 3, we obtain the longevity improvement rates as

$$\begin{bmatrix} {}^{1996}ie_0 & {}^{1997}ie_0 & {}^{1998}ie_0 & {}^{1999}ie_0 \\ {}^{1996}ie_1 & {}^{1997}ie_1 & {}^{1998}ie_1 & {}^{1999}ie_1 \\ {}^{1996}ie_2 & {}^{1997}ie_2 & {}^{1998}ie_2 & {}^{1999}ie_2 \\ {}^{1996}ie_3 & {}^{1997}ie_3 & {}^{1998}ie_3 & {}^{1999}ie_3 \\ {}^{1996}ie_4 & {}^{1997}ie_4 & {}^{1998}ie_4 & {}^{1999}ie_4 \end{bmatrix} = \begin{bmatrix} 0 & \frac{4}{23} & \frac{2}{27} & 0 \\ 0 & \frac{6}{29} & \frac{3}{35} & 0 \\ 0 & \frac{5}{16} & \frac{2}{21} & 0 \\ 0 & \frac{3}{7} & \frac{1}{5} & -\frac{1}{12} \\ 0 & \frac{1}{2} & \frac{1}{3} & -\frac{1}{4} \end{bmatrix} \tag{3.14}$$

Under the previously established criterion, the 1997 cohort clearly is a select cohort with respect to longevity. ◇

Note, that in this case, select cohorts with respect to mortality improvement and with respect to longevity improvement do coincide. But this need not be the case.

To see possible differences between these two criteria, consider another population of guinea pigs with a shorter lifespan (MacMinn et al., 2005).

Example 3 (Population of guinea pigs, comparison of criteria)
Assume three hypothetical generations of three guinea pigs born in years 2000, 2001, and 2002, respectively, with ages at death assumed to be

2000 *birth cohort*	0.50	1.00	2.50
2001 *birth cohort*	0.25	2.00	2.25
2002 *birth cohort*	0.50	1.00	2.50

This data allows us to obtain the central death rates and, subsequently, mortality improvement rates as

$$\begin{bmatrix} {}^{2000}m_0 & {}^{2001}m_0 & {}^{2002}m_0 \\ {}^{2000}m_1 & {}^{2001}m_1 & {}^{2002}m_1 \\ {}^{2000}m_2 & {}^{2001}m_2 & {}^{2002}m_2 \end{bmatrix} = \begin{bmatrix} \frac{2}{5} & \frac{4}{9} & \frac{2}{5} \\ 1 & \frac{1}{2} & 1 \\ 2 & 4 & 2 \end{bmatrix} \tag{3.15}$$

$$\begin{bmatrix} {}^{2001}im_0 & {}^{2002}im_0 \\ {}^{2001}im_1 & {}^{2002}im_1 \\ {}^{2001}im_2 & {}^{2002}im_2 \end{bmatrix} = \begin{bmatrix} -\frac{1}{9} & \frac{1}{10} \\ \frac{1}{2} & -1 \\ -1 & \frac{1}{2} \end{bmatrix} \tag{3.16}$$

The complete life expectancies and longevity improvement rates are given by

$$\begin{bmatrix} {}^{2000}\mathring{e}_0 & {}^{2001}\mathring{e}_0 & {}^{2002}\mathring{e}_0 \\ {}^{2000}\mathring{e}_1 & {}^{2001}\mathring{e}_1 & {}^{2002}\mathring{e}_1 \\ {}^{2000}\mathring{e}_2 & {}^{2001}\mathring{e}_2 & {}^{2002}\mathring{e}_2 \end{bmatrix} = \begin{bmatrix} 1 & \frac{3}{2} & 1 \\ \frac{3}{4} & \frac{9}{8} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \tag{3.17}$$

$$\begin{bmatrix} {}^{2001}ie_0 & {}^{2002}ie_0 \\ {}^{2001}ie_1 & {}^{2002}ie_1 \\ {}^{2001}ie_2 & {}^{2002}ie_2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{3} \\ \frac{1}{2} & -\frac{1}{3} \\ -\frac{1}{2} & 1 \end{bmatrix} \tag{3.18}$$

Clearly, the cohort born in 2001 improves less often than the cohort born in 2002 with respect to mortality improvement, but with respect to longevity improvement it improves more often. ◇

The two concepts discussed so far are related, They are different ways of assessing changes of longevity and/or mortality across generations. The two previous examples are extreme cases of mortality and somewhat artificial, but they show that differences are possible. However, this is not necessarily the case.

In the following chapter, the results of applying both criteria to actual data

are discussed. Interestingly, in some cases, both criteria coincide in determining a cohort as select, but more often this is not the case.

In the context of the previously given example, it may seem reasonable to comment on the use and practical relevance of choosing longevity improvement rates as a criterion for determining select cohorts. While in an example this seems to be a useful approach, it is a lot less relevant in practical applications.

Empirical data found e.g. in life tables is almost always discrete and of annual nature (Bowers et al., 1997). Therefore, it is often not possible to use complete life expectancies. Instead, curtate life expectancies, ${}^{z}e_x$, would have to be used unless there were monthly or even weekly life tables available. But this is almost never the case.

For the calculations performed for various countries, the underlying data was taken from the Human Mortality Database (HMD). This collection of data provides reasonable estimates for complete life expectancies ${}^{z}\mathring{e}_x$ so that it was possible to apply Definition 4, and select cohorts with respect to longevity improvement could be identified.

A Global Criterion

The two criteria introduced in Definitions 2 and 4 can be considered local criteria, since they only account for differences between a select cohort and its neighboring cohorts. This clearly ignores the overall development of mortality improvement (or longevity improvement) over a larger number of birth years.

A way to define a more global criterion is discussed in MacMinn et al. (2005). At each age, a cohort's mortality improvement rates are compared to all other cohorts' mortality improvement rates rather than only to those of the adjacent

cohorts.

Definition 5 (Select cohort with respect to the percentile criterion)

Denote by $\xi^x_{k/100} \quad (x = 0, ..., \omega)$ *the* k*-th percentile of mortality improvements rates at age* x *over all cohorts considered.*

If, for a certain cohort born in year z*, mortality improvement rates of at least* $p\%$ *of ages are higher than* $\xi_{k/100}$*, i.e.*

$$\frac{\#\{x|{}^z im_x \geq \xi^x_{k/100}\}}{\#\{available \;\; {}^z im_x\}} \geq p\%, \tag{3.19}$$

we say that the cohort born in z *is a* ***select cohort with respect to the*** k***-th percentile at*** $p\%$*.* ◇

Considering Longer than Annual Periods

Until now, we only considered annual indices. To incorporate some smoothing of pure mortality data, we will also consider 5-year central death rates and define improvement rates and a criterion for select cohorts analogously to the definitions presented before.

Definition 6 (5-year central death rates)

Recall definition (2.27)*. With* $n = 5$*, and introducing prefices to denote the birth year* z*, we define the* ***5-year central death rate*** *for the cohort born in year* z *as*

$${}^z_5 m_x = \frac{{}^z_5 d_x}{{}^z_5 L_x} \tag{3.20}$$

◇

Definition 7 (5-year mortality improvement rates, MacMinn et al. (2005))

Similarly to Definition 1, we set

$$\begin{aligned} {}^{z}_{5}im_x &= \frac{{}^{z-1}_{5}m_x - {}^{z}_{5}m_x}{{}^{z-1}_{5}m_x} \\ &= 1 - \frac{{}^{z}_{5}m_x}{{}^{z-1}_{5}m_x} \end{aligned} \tag{3.21}$$

◇

Definition 8 (Select cohort with respect to 5-year mortality improvement, MacMinn et al. (2005))

If, for a certain cohort born in year z, at least half of 5-year mortality improvement rates are higher than the corresponding values of the two "adjacent" cohorts born in years $z-5$ and $z+5$, respectively, i.e.,

$$\frac{\#\{x|{}^{z-5}_{5}im_x \leq {}^{z}_{5}im_x \geq {}^{z+5}_{5}im_x\}}{\#\{available\ \ {}^{z}_{5}im_x\}} \geq \frac{1}{2}, \tag{3.22}$$

we say that the cohort born in z is a ***select cohort with respect to 5-year mortality improvement****.*

If the ratio in (3.22) *is at least $p\%$, we say that it is a* ***select cohort at*** *$p\%$.* ◇

Some results of applying this criterion to actual data are shown in the following chapter. However, the increase of longevity - or decline of mortality for a specific cohort does not seem to be accounted for well enough; this criterion identifies less select cohorts than other criteria. Subsequently, emphasis is placed on criteria for annual data.

An Approach from a Monetary Point of View

While all of the previously presented criteria and definitions are used to compare data or indices that are more or less of a demographic nature, we may also consider numbers which directly describe the monetary impact of increased longevity.

Since the price of an annuity is a natural number related to the cost of increased longevity, it seems logical to compare prices for a specific type of annuity with a fixed set of characteristics (such as term period, beginning age, annual payment amounts etc.).

Of course, there is no unique set of characteristics and it is arguable for instance about what interest rate should be used for the calculation. For simplicity, we will start with a fixed reasonable rate such as 3%. The results of taking historical real or nominal interest rates is presented in the following chapter.

Still, we need to define a criterion for a select cohort with respect to this index. As an increase in longevity will be reflected in a higher price for an annuity (the annuity has to be paid over a longer period), it is most reasonable to identify cohorts whose annuity prices are local maxima as select cohorts.

Reflecting the special concerns of public pension systems and private insurance companies providing life annuities, it also seems most natural to take into account prices for annuities over a typical retirement period. For most countries, retirement typically starts around an age of 65 years. Here, the period is limited to 35 years, i.e. to age 100, since there are actually very few observed lives beyond that age.

With these considerations in mind another criterion can be provided (MacMinn et al., 2005). The technical characteristics of the annuities are

understood to remain the same.

Definition 9 (Select cohort with respect to retirement annuity prices)
If, for a certain cohort born in year z, the price of a 35-year annuity starting at age 65 is higher than the prices for the two adjacent cohorts, i.e.

$$ {}^{z-1}\ddot{a}_{65:\overline{35}|} < {}^{z}\ddot{a}_{65:\overline{35}|} > {}^{z+1}\ddot{a}_{65:\overline{35}|} \quad or \quad {}^{z-1}a_{65:\overline{35}|} < {}^{z}a_{65:\overline{35}|} > {}^{z+1}a_{65:\overline{35}|}, \tag{3.23} $$

we say that the cohort born in year z is a ***select cohort with respect to retirement annuity prices****.*
Note that the two expressions in (3.23) *are equivalent; cf.* (2.50)*.*

CHAPTER IV

OVERVIEW OF SELECTED COUNTRIES

Underlying Data and Related Problems

In this chapter, we will investigate the application of a variety of previously presented criteria for select cohorts to actual data. Inspired by the investigations performed by Willets (2004) for England and Wales as well as for Japan, but with a more comprehensive and mathematically more precise set of criteria, results for different countries are presented in detail and general observations are made.

The mortality data used for all calculations in this thesis is obtained from the Human Mortality Database (HMD). This online database is a joint project of the University of California at Berkeley, USA, and the Max-Planck-Institute for Demographic Research in Rostock, Germany. While it provides a variety of demographic data series for a large number of countries, in some cases the practical use is limited due to relatively short periods of data coverage.

For instance, data for Germany is available as three distinct sets: for West Germany and East Germany, and for the reunited Germany starting in 1991. Of course, these data sets cannot be combined, and detailed results for the two former countries are not of special interest anymore. Also, data for countries in Eastern Europe was mostly not available for periods longer than 30 years, and this obviously limits its use.

Nonetheless, the insurance industry in Germany (as well as in other countries with an equally small range of data in the HMD) is of notable size, and there is

reasonable public interest in investigating possible cohort effects. For this thesis only a limited number of countries was investigated. Further research in this field is necessary, though, and therefore recommended.

The range of ages given by the data in the HMD is 0 through 109 years and there is always one additional entry for 110 years and beyond, denoted by "110+". We will adopt this notation and assume a unique age of 110 years for this particular group. This is a simplification, but due to the extremely small number of observed lives beyond 110 years it should not have a notable effect.

For the calculations performed, period life tables were taken. This means that mortality data is given by the year of death and the corresponding year of birth was calculated.

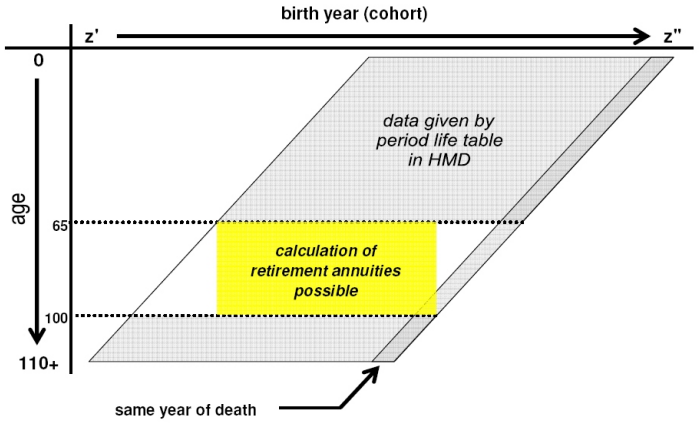
It is a consequence of using period life tables that for the very first year z' in each collection, there are few individuals who died at age 110+ and their birth year was 110 years earlier, i.e. in calendar year $z' - 110$.

Consequently, the earliest possible year for calculations is $z' - 110$ with 110+ being the only age for which mortality data is available for that cohort. For year $z' - 109$, there is data at ages 109 and 110+ and so on. On the other hand, for the most recent year z" in the HMD, we only have data at age 0. This fact should be kept in mind when looking at the evaluation of data series since we will say that data is available for calendar years (or cohorts) $z' - 110$ through z''.

Figure 8 gives a schematic view of the data series as it can be obtained from the HMD. The horizontal axis represents birth years (cohorts) while ages are represented on the vertical axis. Since data in period life tables is given by year of death, each such set of data corresponds to a sequence of numbers in a "northeast-southwest direction", and the given data points cover a

parallelogram-shaped area.

Figure 8
Schematic view of data available from the HMD



When calculating retirement annuities, $a_{65:\overline{35}|}$, i.e. annuities beginning at age 65 over a period of 35 years, another restriction is due to this situation. The arrangement of available data translates into not being able to perform these calculations for cohorts born later than $z'' - 100$. The range for which these annuities can be obtained is highlighted in Figure 8.

Taking the data provided by the Human Mortality Database only permits calculation of retirement annuity prices only for cohorts born before around 1900. This, of course, is dissatisfactory because these earlier cohorts are not of critical importance for retirement systems and pension providers any more, and more recent cohorts are more interesting.

The solution to overcome this flaw would be to project incomplete mortality data for younger cohorts, born in e.g. year $z'' - 50$. Under this approach, prices of

retirement annuities for cohorts that have not yet reached their retirement age would be estimated or predicted. However, to have sufficiently stable projections of mortality data suitable models have to be used and relevant parameters must be estimated in the process of model-fitting. But these considerations are beyond the scope of this thesis. Nonetheless they could be helpful to base observations on a criterion that has a different point of view than those considering pure mortality data such as mortality or longevity improvement.

To clarify and better understand the results that will be presented below, it should be mentioned here that all data series in the HMD are available for the male and the female population (as is commonly the case with all mortality-related data in actuarial mathematics), but also for the overall population on a combined basis.

Observations are therefore separately described for "Males", "Females", and the "Combined" population. When referring to these three collections of data, we will also name them "the three data sets or series".

Detailed Results by Country

Results of applying different criteria are shown and discussed in detail by country in an alphabetical order. Note that data for "England and Wales" does not reflect the overall situation for the United Kingdom, but is provided this way by the HMD. Therefore, it is treated as one country for the purpose of this thesis only.

Canada

Mortality data for Canada was available for calendar years 1812 through 1996.

Neglecting very early as well as the most recent years which provide only a

few number of ages, most criteria produce a relatively small number of select cohorts. With respect to mortality improvement at 60%, the cohort born in 1902 is identified as a select cohort - regardless of whether it is the male, the female, or the overall population. However, with respect to longevity improvement at 60%, there are no select cohorts.

Table 2
Select cohorts (5-year mortality improvement at 60%; Canada)

Males	Females	Combined
	1872	
1895		1895
	1899	
1905	1905	1905
1906	1906	1906
1919	1919	1919
1944		

Data Source: Human Mortality Database (2005); own calculation

Table 3
Select cohorts (60-th percentile at 60%; Canada)

Males	Females	Combined
1891		1891
1902	1902	1902
		1916
1919		1919
1934		1934
	1948	1948

Data Source: Human Mortality Database (2005); own calculation

Table 2 shows select cohorts with respect to 5-year mortality improvement at 60%. Mostly, these select cohorts coincide for the three data sets. This is not the case when looking at select cohorts with respect to the 60-th percentile at 60%, the data being shown in Table 3. The only year that appears for both criteria is 1919.

The criterion for retirement annuity prices does not identify any cohort born later than 1900 as select for the reasons explained above. However, during the 1880's there are numerous select cohorts, mainly among males.

England and Wales

England and Wales is one of the countries with the largest data collection available from the Human Mortality Database, i.e. for the period 1733-1998.

The situation here is similar to what it was for Canada. Again, there are no select cohorts identified with respect to longevity improvement and only a small number with respect to mortality improvement that is consistent across the data sets. These cohorts are listed in Table 4. They partly include those select cohort years that Willets (2004) identified - with the exception of male cohorts centered around 1931.

Table 4
Select cohorts (mortality improvement at 60%; England & Wales)

Males	Females	Combined
1919		1919
1921	1921	1921
1946	1946	1946
1948		1948

Data Source: Human Mortality Database (2005); own calculation

The 5-year mortality improvement criterion identified mostly cohorts born before 1900 as select. However, 1919 is a select cohort for females as well as for the combined population as is the case with 1946 for the male and the combined population - all at a 60% threshold.

Surprisingly, under the percentile criterion (with respect to the 60-th percentile at 60%) a large number of male, female, and combined select cohorts

coincide, namely 1918, 1919, 1921, 1932, 1946, and 1948.

Similar to the Canadian data, earlier select cohorts (before 1900) with respect to retirement annuity prices are quite uniformly spread and occur for the three data collections equally often; within each decade there are two or three select cohorts within each of the data sets.

Finland

The results that can be obtained from the Finnish data are somewhat similar to what could be observed for Canada and England and Wales. The range of years (1769 through 2002) has a length in between the two previously discussed countries. In contrast, the select cohorts are different for the three data sets (males, females, and combined population) throughout all criteria but the annuity price criterion.

With respect to mortality improvement at 60%, the 1836, 1919, and 1921 male cohorts are identified. The latter two cohorts are select for the combined population as well. There are no female select cohorts after 1900, but 1947 is also select for the overall population.

Again, we have no select cohorts under the longevity improvement criterion and there is only one select cohort under the 5-year mortality improvement criterion (1945 for the male and the combined population). Contrarily, there is a larger number of select cohorts under the percentile criterion, listed in Table 5.

The situation with respect to retirement annuity prices is the same as before: a relatively large number of select cohorts is identified among the generations born in the second half of the nineteenth century. They are quite uniformly spread; no specific accumulation can be identified.

Table 5
Select cohorts (60-th percentile at 60%; Finland)

Males	Females	Combined
	1809	
		1814
		1820
		1822
		1832
1836		1836
1869	1869	1869
1919		1919
		1921
1925		1925
	1941	1941
		1945
		1947
	1965	
	1969	
	1971	

Data Source: Human Mortality Database (2005); own calculation

France

The range of available data for France is about the average of all countries, i.e. 1790 through 2002. Observations made upon application of the established set of criteria to the data fit into the previously described situation. No select cohorts with respect to longevity improvement, and only a small number with respect to mortality improvement (annual as well as 5-year improvement rates) are identified.

Select cohorts with respect to mortality improvement are listed in Table 6. With respect to the 5-year criterion, cohorts born in 1916 and 1925 - regardless of whether looking at males, females or the combined population - are select.

The particular observation that cohorts appear simultaneously in all three data sets can also be made with respect to the 60-th percentile at 60%, see Table 7.

In contrast to the relatively large number of select cohorts with respect to

Table 6
Select cohorts (mortality improvement at 60%; France)

Males	Females	Combined
		1848
1916	1916	1916
1919	1919	1919
1921	1921	1921
1941		1941

Data Source: Human Mortality Database (2005); own calculation

Table 7
Select cohorts (60-th percentile at 60%; France)

Males	Females	Combined
	1848	1848
1850		
1872		
1916	1916	1916
1919	1919	1919
1921	1921	1921
1941	1941	1941
1972		
1976		1976

Data Source: Human Mortality Database (2005); own calculation

retirement annuity prices (at 60%) that were previously identified, the French data exhibits a much smaller number, only about two select cohorts per decade in the nineteenth century. As before, most cohorts are select regardless of which data set (males, females, and combined population) is considered.

Italy

Italy is one of the countries for which the Human Mortality Database provides the largest data series, the range being 1763 through 2001. The number of select cohorts under each of the criteria shows similar patterns as with all of the previously discussed countries. However, select cohorts do not coincide in the three data series to the extent as before.

Under the mortality improvement criterion, only 1919 and 1921 stand out as select cohorts for both males and females as well as for the overall population. Few other select cohorts are identified in the first half of the twentieth century.

Again, no cohorts are select with respect to longevity improvement. But in contrast to e.g. the French data, there is a large number of select cohorts with respect to 5-year mortality improvement, mostly for the nineteenth century and coinciding across the data sets. The most recent select cohorts are 1949 (males), 1925 and 1939 (females), and 1939 and 1949 (combined).

The percentile criterion identifies 1818, 1917, 1919, 1921 and 1923 as select - uniformly in all three data sets. Furthermore, there are select cohorts found in 1929, 1935 and 1945 (males), in 1910, 1912, 1914, 1937 and 1942 (females), and in 1929, 1932, 1935, 1937, 1939, 1942 and 1945 (for the combined series).

Analyzing annuity prices (retirement annuities at 60%), the select cohorts are mostly congruent across the data sets with frequencies of about three to four per

decade.

Japan

The frequency and distribution of select cohorts is slightly different for Japan. Besides, Japan is the country with the second-smallest range of mortality data available, providing series for cohorts 1841 through 1999.

In addition to the longevity improvement criterion not producing any select cohorts, the annuity price criterion does not identify any either. Contrarily, there is quite a large number with respect to the mortality improvement criteria (both annual and 5-year data), and even more under the percentile criterion. The select cohorts are given by Tables 8, 9 and 10, respectively.

Table 8
Select cohorts (mortality improvement at 60%; Japan)

Males	Females	Combined
	1888	
	1897	
1899	1899	1899
1906		1906
1908		1908
		1913
	1919	1919
1921		1921
1930		1930
	1934	1934
		1937
1939	1939	1939
1943	1943	1943
1946		1946
	1948	
1966		1966
1970		

Data Source: Human Mortality Database (2005); own calculation

Table 9
Select cohorts (5-year mortality improvement at 60%; Japan)

Males	Females	Combined
	1912	1912
	1913	1913
	1925	
1937		
1939		1939
1940		1940
1946	1946	1946
1949		1949

Data Source: Human Mortality Database (2005); own calculation

Table 10
Select cohorts (60-th percentile at 60%; Japan)

Males	Females	Combined
1888	1888	1888
	1891	1891
	1897	1897
1899	1899	1899
1906		1906
1908	1908	1908
1913	1913	1913
	1917	
	1919	1919
1921		1921
1934	1934	1934
1937		1937
1939		1939
1943	1943	1943
1945	1945	1945
1946	1946	1946
1948		

Data Source: Human Mortality Database (2005); own calculation

Only under the percentile criterion do male and female select cohorts coincide to a significant extent. It is interesting to observe that so far Japan is the first country for which select cohorts seem to exist beyond 1950 (neglecting of course very recently born cohorts; cf. the above remarks).

Netherlands

For the countries discussed so far, there is reasonable justification to claim that select cohorts at least exist. Even if not all criteria coincide in the select cohorts they identify, most give numerous candidate cohorts. This is not the case for other countries, the Netherlands being the first country discussed here.

Neither of the criteria based on the comparison of indices from mortality and longevity data give more than just one or two select cohorts - although the data range (1741 through 1995) is one of the largest available in the Human Mortality Database.

However, when investigating retirement annuity prices, there is a significantly long list of cohorts identified as select (at 60%). About half of these select cohort years coincide across the male, the female and the combined population. These are mainly in the 1810s, in 1838 and 1840, around 1850 and during the 1860s. Non-congruent select cohorts are spread all over the ninetenth century.

Norway

Data for Norway is provided for 1737 through 1991, which is on the upper end with respect to years covered. In spite of a large data range, it is one of the countries with the smallest overall number of select cohorts - supporting the notion that there is no significant cohort effect in Norway.

Criteria directly based on mortality and longevity did not identify many select cohorts; among the few select cohorts is the female cohort born in 1816, with respect to mortality improvement as well as to the 60-th percentile (both at 60%).

Until about 1850, almost all select cohorts with respect to annuity prices coincide in all data series. This is not the case anymore for the years until around 1900.

Sweden

For Sweden, the Human Mortality Database provides mortality data as early as 1647, but due to limitations arisen in the process of data analysis, only years 1742 and thereafter have been considered.

Select cohort identification follows the same pattern as in the Netherlands. Apart from 1921 being a select male and combined cohort year under the (annual) mortality improvement criterion at 60% as well as 1903 being select for females and the overall population with respect to 5-year mortality improvement at 60%, there are only a few cohorts identified as select under the percentile criterion. These are listed in Table 11.

Table 11
Select cohorts (60-th percentile at 60%; Sweden)

Males	Females	Combined
1921	1921	1921
1923		
		1928
1962		1962
	1964	
1965		
	1968	
1969		

Data Source: Human Mortality Database (2005); own calculation

In contrast to this rareness of select cohorts, the annuity price criterion produces the largest number of select cohorts among all countries. Almost uniformly, every second or third generation in the nineteenth century is select, mostly coinciding across the male, the female and the combined data series.

The overall situation could be a hint that there is no pronounced cohort effect present. A certain unevenness which can be seen in the series of annuity prices can be considered normal. The uniformity with which local maxima are spread over time (and consequently, also local minima) shows that there is no specific accumulation that could be observed for other countries.

Switzerland

Swiss data encompasses the cohorts born in 1767 through 2002; this is a medium range within the countries considered. With respect to the mortality improvement criterion, only the male and combined cohorts of 1818 are select. The longevity improvement criterion does not produce any reasonable select cohorts (cf. to the above remarks concerning the number of ages available for calculations).

Contrarily, with respect to both 5-year mortality improvement and the 60-th percentile at 60%, there are more select cohorts. While the former approach gives three select cohorts uniformly for all three series (1848, 1855, and 1871 - but none in the twentieth century), there is only one uniform select cohort (1818) and a small number of scattered other with respect to the latter criterion.

Cohorts identified as select with respect to retirement annuity prices are listed in Table 12. They are mostly uniform across the data sets, especially during the first half of the nineteenth century.

Table 12
Select cohorts (retirement annuity prices at 60%; Switzerland)

Males	Females	Combined
1818	1818	1818
	1822	1822
1823		
1825		1825
	1826	
1829	1829	1829
1833	1833	1833
1836		1836
1838	1838	1838
1844	1844	1844
1848	1848	1848
1850		
1855	1855	1855
	1858	
1859		1859
1862	1862	1862
1865	1865	1865
1872		1872
	1873	
1877	1877	1877
1881		
	1882	
1883		1883
	1885	
1887		
1891		
1893		
1896		

Data Source: Human Mortality Database (2005); own calculation

United States

Among those countries that were investigated for the purpose of this thesis, the United States have the smallest range of data available from the HMD with cohorts of 1850 through 1999. This causes the number of identified cohorts to be relatively small.

Given the restricted range of data, retirement annuity prices could only be calculated for cohorts 1894-1899. Subsequently, it is not justified to apply the select cohort criterion for this set of data. Also, as could be seen for all other countries, there are no select cohorts with respect to longevity improvement.

However, there is an unusually long list of cohorts that have be selected with respect to mortality improvement, and most of them are congruently select across the three data sets. The complete list is given by Table 13.

Table 13
Select cohorts (mortality improvement at 60%; United States)

Males	Females	Combined
1899	1899	1899
1902	1902	1902
1909	1909	1909
1912	1912	1912
1919	1919	1919
1921	1921	1921
1929		1929
		1931
1934	1934	1934
1942		1942
1944	1944	1944
1946	1946	1946
1948	1948	1948

Data Source: Human Mortality Database (2005); own calculation

While the 5-year mortality improvement rates produced only a small number of select cohorts (namely 1919 for the male and combined population, and 1938 jointly for the three data sets), there is a much larger number obtained through the percentile criterion. Table 14 identifies these cohorts.

Table 14
Select cohorts (60-th percentile at 60%; United States)

Males	Females	Combined
		1871
1881	1881	1881
1887		
	1891	1891
		1894
1899	1899	1899
1902	1902	1902
1909	1909	1909
1912	1912	1912
1919		1919
1921	1921	1921
1929		1929
	1931	
1934	1934	1934
1937	1937	1937
1942		1942
	1944	
1946	1946	1946
	1948	1948

Data Source: Human Mortality Database (2005); own calculation

CHAPTER V

OBSERVATIONS AND ANALYSIS OF RESULTS

Classification into Groups of Countries

Focusing less on particular select cohorts, but rather comparing the number identified by each of the criteria presented in the previous chapter, three different patterns may be distinguished:

- The first group of countries follows a pattern that can be characterized by a very small number of select cohorts identified under the (annual and 5-year) mortality improvement criteria as well as a small to moderate number with respect to the 60-th percentile at 60%.

  However, the largest such list of select cohorts can be found for select cohorts with respect to retirement annuities at 60%.

  An example of a country exhibiting this pattern is Sweden. Although it is a country with one of the largest sets of data, there is only a very small number of select cohorts identified by the mortality-based and longevity-based criteria. On the other hand, the annuity price criterion produces a very large number of select cohorts. This is, of course, partly due to the enormous range of years provided by the Human Mortality Database.

  Having a look at the list of select cohorts leads to the conjecture that there is no such select cohort effect as it was identified by Willets for other countries.

  When considering uneven development of annuity prices, the list of generations, whose annuity prices are local maxima, (i.e., they are higher

than the corresponding prices for the generations born one year before and one year after) gives a very uniformly spread list of cohorts, which could be some evidence that there is no - at least no pronounced - cohort effect.

These different observations together support the claim that the Swedish population does not exhibit any select cohort patterns.
The Netherlands is the other country for which results follow the pattern described above. Besides, a range of data can be obtained from the HMD that is equally extended as the Swedish data.

The percentile criterion identifies even less cohorts as select, and there are also less select cohorts with respect to retirement annuity prices. Another slight difference compared to Sweden is the fact that under the latter criterion select cohorts do not coincide across the male, the female and the combined population as much.

- Different patterns can be observed for Japan and the United States, the two countries with the smallest available data range in the HMD. This limited availability of data makes it next to impossible to apply the annuity price criterion. Consequently, for these two countries there are no cohorts that have been identified as select with respect to this criterion.

  A difference to Sweden and the Netherlands is common to the two countries. It consists in a relatively large number of select cohorts under the percentile criterion. For both, there are roughly ten to fifteen select cohorts for each of the three data sets. About two thirds of these cohorts do in fact coincide - a peculiarity which cannot be identified to this extent for the other countries. In particular, for Japan there are only three select cohorts with respect to

mortality improvement that are congruent across data sets in addition to five to ten individual select cohorts.

When looking at 5-year rather than at annual improvement rates, only one select cohort (1946) is common to all three series. The number of individual cohorts given by this criterion is also smaller than in the case of the annual criterion.

For the United States, similar observations can be made. The 5-year mortality improvement criterion also gives only one consistent select cohort (born in 1938). Yet, with the exception of two (for males) or three cohorts (for the combined series), all select cohorts with respect to the corresponding annual improvement rates coincide across the three data sets.

Also, the majority of select cohorts under the percentile criterion is consistent, with only a few additional years that are specific to a single data set.

- The vast majority of countries investigated for this thesis belongs to the third group. Its pattern can be characterized by a very small number of select cohorts with respect to (annual) mortality improvement. As a general rule, this is also the case for the 5-year mortality improvement rates.

  The retirement annuity price criterion produces almost always a sizable list of select cohorts - provided sufficient data is available. The list then comprises between eight and 27 generations. Only Canada is an exception with only six, one, and two years, respectively, the cohort born in 1887 being the only common select cohort. In general, the proportion of consistent select cohorts across data series is quite high, almost always above two thirds.

  Norway is probably the country with the smallest number of select cohorts

while at the same time the Human Mortality Database provides one of the most extended data sets for this country. The mortality improvement and the percentile criterion producing only one and four select cohorts, respectively, is the main specialty of Norwegian data within this third group of countries. Two other countries also diverge from the general pattern. Data for Italy as well as for England and Wales produces unusually large numbers of select cohorts under the 5-year mortality improvement and the percentile criterion. Their number is between ten and twenty years with a high proportion of cohorts coinciding across the data sets.

Observation of Common Patterns

All countries have in common that there are no select cohorts with respect to longevity improvement at 60%. A possible alteration of this particular criterion would consist in lowering the threshold percentage, thus making it easier for cohorts to be identified as select by this criterion.

However, in most cases such a lowering had to be performed to a threshold as low as 40% in order to have a list of select cohorts of more than just ten years. For instance, in the case of England and Wales, lowering this threshold even further to 35% produced an enormous list of select cohorts comprising all but very few years. Clearly, such a low percentage does not make any sense if at the same this translates into identifying almost all generations as select cohorts.

Similar behavior can be observed for the Norwegian data set. Only when lowered to 40%, there was a notable number of select cohorts with respect to longevity improvement identified. But this list included e.g. for females, only 1904

and then every third or fourth cohort beyond 1950. Again, this seems to be an unreasonable results, possibly including too many cohorts which actually might not be "special" enough in terms of their mortailty improvement to justify them being identified as select.

A possible explanation for this observation could be the fact that mortality data and especially its changes over time are being too much compressed when taking longevity improvements. Subtle changes that would be identified by the analogously defined mortality improvement criterion are not detected anymore. Hence, the evaluation of results under this specific criterion requires special awareness.

Also, it is interesting to observe that those cohorts in England and Wales as well as in Japan that had been identified as select cohorts by Willets (2004) could not be as clearly identified as it might be expected. Only under few criteria do these specific cohorts appear in the tables given above.

But this does not mean that Willets' observations are per se wrong or that his criteria give misleading results. Neither does it imply that the criteria described in this thesis are useless.

A reason for this divergence might rather be the fact that Willets performed the smoothing of raw mortality data described in the beginning of the previous chapter. As could also be seen (cf. figures in Chapter III), conceptual changes (such as narrowing the smoothing period or redefining the threshold percentage) caused the results to change notably. Although these alterations were only shown for the specific data of England and Wales, the outcomes are similar for other countries.

Analysis and Appraisal of Criteria

With the definitions established in Chapter III, only qualitative aspects of cohorts being select or not are captured. It would require further analysis of the underlying data and the establishment of a new set of definitions and criteria to actually measure the quantitative aspects of uneven mortality improvement, thus being able to rank select cohorts by "amounts of improvement".

Especially the results produced by the annuity price criterion make it seem dissatisfactory. Not only regarding its inability to operate on the incomplete data series beyond about 1900, but also with respect to the fact that by identifying those cohorts as select for which annuity prices have local maxima, the nature of discontinuity obviously is not fully capture since it neglects the overall pattern of change.

As the 5-year mortality improvement rates were calculated under the assumption of a uniform distribution of deaths within the year of death (UDD assumption; cf. Chapter III, (2.31), p. 14), the question is to what extent the select cohorts with respect to 5-year mortality improvement that have been identified in the preceding investigation are dependent of this specific assumption. However, the way chosen in this thesis was the most reasonable method and, although the UDD assumption can be termed somewhat theoretical, it is not too unreasonable to use it as a rough approach in order to analyze mortality improvement.

A remark on the usefulness of the longevity improvement criterion also seems to be indicated here. While it is defined in a similar fashion as the mortality improvement criterion, it obviously is much less useful. The only reason for this can be seen in the set of indices that is analyzed by this criterion: longevity improvement rates at age x for the generation born in year z, ${}^{z}ie_{x}$. In contrast to

mortality improvement rates which seem to be able to capture much of the uneven nature of mortality over a cohorts' life, this is not the case with the longevity improvement rates.

Consequently, this criterion must be deemed less useful as the other criteria, especially those related to mortality improvement rates.

Further Questions

So far, we have studied possible measures for mortality improvement and analyzed select cohort patterns for various countries. At this point, the question for the reason of the existence of select cohorts may arise. A clear answer, however, is beyond the scope of this thesis.

Numerous investigations have already been performed on this issue, with much of the interest arisen from a medical background.

Willets (1999) shows trends of mortality for a number of causes of death. But he also states that for England and "the majority of deaths for younger males are non-health-related - i.e. violent or accidental", the latter reason accounting for 62% of deaths for men aged 20 to 29. The "'big three' killers - i.e. cancer, heart disease and stroke" which are the main causes of death related to diseases account for only 10%.

For the United States, the situation is somewhat the same. The National Center for Injury Prevention and Control (2005) states heart diseases as the main cause of death for the overall population, but at the same time "unintentional injury" is the undisputed leading cause of death for males under 44 years as well for females younger than 34 - but in both cases beyond age 1.

In Willets (2004), further investigation of these specific causes of death are presented. Also, possible explanations are sought in the behavior of certain generations and public consciousness towards health-related issues.

Once more bearing in mind the specific concerns of pension providers, the primary question rather is: "Is there a cohort effect or not and how strong is it?", for this should have an effect on reserving, funding and pricing policies.

A secondary issue in this context is the question for the exact reasons which cause certain cohorts to be so much outstanding versus other cohorts such that the above observations can be made. Demographers' and public interest in an answer to this question clearly exists, but a pension provider - be it a governmental agency or a private life insurance company - will most likely be less concerned with specific reasons.

This is the main reason why this thesis concentrates on the identification of select cohorts.

CHAPTER VI

SUMMARY AND CONCLUSIONS

So far, this thesis has provided an overview of concepts and notations from actuarial mathematics to describe mortality and longevity. Also, the calculation of basic types of life annuities has been explained.

Starting from the notion of the "cohort effect" that Willets (1999) identified for the population of the United Kingdom, his criterion for identifying "select cohorts", i.e. generations with more rapid improvement of mortality, has been analyzed.

The results were compared to those that were obtained through possible changes of this very first criterion. It could be observed by means of graphical representations that those cohorts originally identified by Willets gradually disappeared. This gave rise to the question for alternative criteria that are mathematically more precise compared to Willets' graphical comparison.

Based on the mathematical concepts introduced in Chapter II, different measures of changes in mortality and longevity have been defined in Chapter III as well as a set of criteria to identify select cohorts. This approach closely follows MacMinn et al. (2005).

These criteria have then been applied to real mortality data from the online Human Mortality Database for a total of eleven countries. A detailed description of the results for each country is given in Chapter IV while a more general overview and a classification of patterns into three groups of countries can be found in

The first such group, comprising Sweden and the Netherlands, is characterized by a very small number of select cohorts under mortality-based and longevity-based criteria, in spite of having the most extended data sets. On the other hand, the annuity price criterion gave a very large number of select cohorts that was uniformly spread over decades. Taking the results of the different criteria together, it can be concluded that there are no (pronounced) cohort effects present in these countries.

Japan and the United States are the two countries for which the smallest data sets were available. This made it almost impossible to apply the annuity price criterion. Contrarily, especially the percentile criterion and the annual mortality improvement criterion produced numerous select cohorts, both criteria mostly coinciding in the years identified.

This was not the case for the majority of countries. While the annuity price criterion always gave the largest number of select cohorts, these were all born in the nineteenth century which is due to a conceptual weakness of this criterion. The number of select cohorts with respect to the 60-th percentile as well as to 5-year mortality improvement were mid-sized while the least were identified with respect to annual mortality improvement rates.

In all cases, the previously established longevity improvement criterion did not identify any select cohorts unless the threshold percentage was lowered to about 40%. But at the same time, this gave an enormous list of all but very few generations as could be seen for England and Wales and for Norway. A possible explanation for this flaw could be that longevity improvement rates as defined above compress mortality changes too much to be able to capture subtle changes to

a reasonable extent.

Another observations was that those cohorts identified as select by Willets (2004) could not be verified to be select under each criterion. This is possibly due to the fact that Willets performed a smoothing process to raw mortality data before computations while all own investigations for this thesis were based on raw data as provided by the Human Mortality Database.

Obviously, there is a another conceptual flaw. Under the annuity price criterion, no cohorts born beyond about 1900 can be investigated. Further analysis of data including projections of mortality to overcome this problem seems necessary and is therefore highly encouraged.

The initial notion of the existence of select cohorts which was claimed by Willets (2004) for England/Wales and Japan could not be unanimously verified for other countries. While in some cases, there is some evidence for their existence, the results for others can be interpreted as not supporting this notion.

However, the intention of this thesis was to further develop Willets' approach and provide possible alternative criteria, but the results obtained did not always clearly identify select cohorts.

Further investigation of this phenomenon and development of more sophisticated measures of mortality and longevity improvement is therefore necessary.

REFERENCES

Arias, Elizabeth. 2004. United States Life Tables, 2002. *National Vital Statistics Report* 53(6). National Center for Health Statistics. Retrieved June 20, 2005 from http://www.cdc.gov/nchs/data/nvsr/nvsr53/nvsr53_06.pdf.

Bowers, Newton L., Hans U. Gerber, James C. Hickman, Donald A. Jones, and Cecil J. Nesbitt. 1997. *Actuarial Mathematics*. 2nd ed. Schaumburg, Illinois: Society of Actuaries.

Human Mortality Database. 2005. University of California, Berkeley (USA) and Max-Planck-Institute for Demographic Research, Rostock (Germany). Available online at http://www.mortality.org on February 20, 2005.

Kinsella, Kevin, and David R. Phillips. 2005. Global Aging: The Challenge of Success. *Population Bulletin* 60(1):4. Population Reference Bureau, Washington, DC. Retrieved June 18, 2005 from http://www.prb.org/Template.cfm?Section=PRB&template=/ContentManagement/ContentDisplay.cfm&ContentID=12430.

MacMinn, Richard D. 2003. International Mortality Comparisons. Presented at the 2003 Vancouver Spring Meeting of the Society of Actuaries, June 23, 2003. Available online at http://journalofriskandinsurance.org/macminn/mortality/ on June 19, 2005.

MacMinn, Richard D., Krzysztof M. Ostaszweski, Ranee Thiagarajah, and Frederik Weber. 2005. An Investigation of Select Birth Cohorts. Presented at the Living to 100 and Beyond Symposium, sponsored by the Society of Actuaries, Orlando, Florida, January 12-14, 2005.

Maier, Heiner, and Rembrandt Scholz. 2004. Immer mehr Menschen können 105. Geburtstag feiern. *Demografische Forschung - Aus erster Hand* 1:4. Available online at http://www.demografische-forschung.org/archiv/defo0401.pdf on June 15, 2005.

National Center for Injury Prevention and Control. 2005. WISQARS Leading Causes of Death Reports, 1999 - 2002. Online database query, retrieved June 19, 2005 from http://webapp.cdc.gov/sasweb/ncipc/leadcaus10.html.

O'Brien, Chris. 2003. The Challenge of Increased Longevity for Pension Providers. Presented to the 25th U.K. Insurance Economists' Conference, Nottingham

University Business School, March 2003. Available online at http://www.nottingham.ac.uk/business/cris/ukec/2003paper.doc on May 15, 2005.

Society of Actuaries. 2003. Report of the Society of Actuaries Mortality Improvement Survey Subcomittee. Retrieved June 15, 2005 from http://www.soa.org/ccm/cms-service/stream/asset/?asset_id=9524077.

Vaupel, J.M., and K.G. von Kistowski. 2005. Der bemerkenswerte Anstieg der Lebenserwartung und sein Einfluss auf die Medizin. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 5:586–592. Available online at http://www.demogr.mpg.de/publications/files/1956_1118841030_1_Bundesgesundheitsblatt48-5.pdf on June 18, 2005.

Wikipedia. 2005. Cohort (statistics) — Wikipedia, the free encyclopedia. Retrieved June 19, 2005 from http://en.wikipedia.org/wiki/Cohort_(statistics).

Willets, Richard C. 1999. Mortality in the Next Millennium. Paper presented to the Staple Inn Actuarial Society. Available online at http://www.sias.org.uk/papers/mortality.pdf on March 7, 2005.

———. 2004. The Cohort Effect: Insights and Explanations. Presented to the Faculty of Actuaries, March 15, 2004, and to the Institute of Actuaries, April 26, 2004. Printed by Bell & Bain Ltd., Glasgow. Available online at http://www.actuaries.org.uk/files/pdf/sessional/sm20040426_cohort.pdf on March 10, 2005.

Yue, Jack C. 2002. Oldest-Old Mortality Rates and the Gompertz Law: A Theoretical and Empirical Study Based on Four Countries. Presented at the Society of Actuaries' Living to 100 and Beyond: Survival at Advanced Ages Conference, Lake Buena Vista, Florida, January 2002. Retrieved June 15, 2005 from http://www.soa.org/ccm/cms-service/stream/asset?asset_id=1215068.