3 Five Hypotheses

The previous chapter explores whether differences in life span after age 50 by month of birth exist in the four populations Austria, Australia, Denmark and Hawaii. This chapter is devoted to testing five hypotheses. The first hypothesis assumes that the interaction between age and the seasons of mortality causes the differences in life span by month of birth. For example, people born in April are older than people born in November when the high mortality of winter strikes them. The second hypothesis tests whether the differences are due to unobserved social factors that influence or result from the seasonal timing of births. The third hypothesis considers whether deadlines for school attendance may have an impact. The fourth hypothesis explains the differences in adult life span by differential survival in the first year of life, while the fifth hypothesis assumes that debilitation *in-utero* or in the first year of life increases the infant's susceptibility to diseases at adult ages.

3.1 Seasonal Distribution of Deaths

The first hypothesis assumes that the observed relationship is caused by the interaction of the seasonal distribution of deaths and the monthly increase in adult mortality. For example, since the April-born are older than the autumn-born when the high mortality risk of winter strikes them they may experience higher mortality overall. In other words, month of birth is not so much related to the beginning or life but is rather a life-cycle indicator. This hypothesis has already been widely discussed and tested in the research about schizophrenia. In this research it is referred to as the "age prevalence and age incidence effect". Torrey et al. (1997) review a large number of studies that deal with this "statistical artifact", and its description here is largely based on Torrey's article. The studies about age prevalence and age incidence deal with the problem of the interaction between age and seasonality in the incidence of schizophrenia, although the term seasonality is not mentioned explicitly. Lewis and Griffin (1981, p. 590) explain the effect as follows:

"For example, people who are 30 years old have a higher incidence of schizophrenia than people who are 29 years old. Therefore, people born in 1937, may have a higher incidence in 1967 than in 1966. When first admissions from 1967 are pooled with first admissions from 1966, people born in 1937 and admitted in 1967 will have more influence on the final result than people born in 1937 and admitted in 1966. The difference in the influence between any 2 years is not great, but over the course of an entire data set it becomes noticeable and should be corrected."

Lewis and Griffin try to explain – in a somewhat complicated manner – that the incidence of schizophrenia is seasonal and that people born in different seasons of the year are hit by the higher risk of developing schizophrenia at slightly different ages. This is similar to the problem one faces with mortality.

Torrey et al. write that although "the age incidence and the age prevalence effects have elicited extensive and spirited discussion among statisticians and psychiatric researchers (...), the effects have been shown to be negligible." Hare (1975) and Lewis and Griffin (1981) had proposed a formula to correct for the effect. Torrey et al. mention thirteen studies that used the correction but still found significant winter–spring birth excess among schizophrenic patients.

In Denmark, Austria, and Australia the risk of death is lowest in the summer and highest in the winter. The differences between the summer trough and the winter peak are considerable, and they depend on the time period, age, climate, and social conditions. In all three countries death rates after age fifty increase by about 0.8% per month of age. The question is whether the increasing mortality with age together with the seasonal differences in mortality may cause the differences in life span by month of birth.

In the following three approaches are used to explore the impact of the seasonal distribution of deaths on the month-of-birth pattern in life span. The first approach is merely descriptive and displays the death rates by month of birth, current month and age in month. The second approach estimates the age-specific force of mortality depending on the month of birth and the current half-year. The third approach models the age-specific death rates by month of birth, current month and age in months.

3.1.1 A Descriptive Approach Based on Death Rates

In the following, a descriptive approach is used for a first analysis of the season of birth effect. On the basis of the Danish longitudinal data, the population at risk and the number of deaths by month of birth and age in months are calculated. Let x denote age in integer years and let y denote the current year. Let i denote age in months since last birthday, j the current month, and k the month of birth; let D_{ijk}^{xy} be the number of deaths that occur for people of age $x + i/12$ years at time $y + j/12$, and let T_{ijk}^{xy} be the corresponding size of the exposed population at risk of death. Note that the month of birth k is already defined by the combination of current month j and age in months since last birthday i. Let n_j denote the number of days in current month j. When calculating the exposed population, all deaths were assumed to occur in the middle of each month. The relative deviation of the monthly death rate from the annual death rate, R_{ijk}^{xy}, was calculated as

$$R_{ijk}^{xy} = \frac{D_{ijk}^{xy}}{T_{ijk}^{xy} n_j} \Big/ M_{xy} \qquad [3.1]$$

where

$$M_{xy} = \frac{\sum_{i=1}^{12} \sum_{j=1}^{12} D_{ij}^{xy}}{\frac{365.25}{12} \sum_{i=1}^{12} \sum_{j=1}^{12} T_{ij}^{xy}}. \qquad [3.2]$$

Let M denote the average death rate D/T over the whole observation period and over all ages. Summing the number of deaths and the population at risk over age x, time y, current month j and age since last birthday i and dividing it by the average death rate M then gives R_k (Equ. 3.3) the death rates by month of birth as compared to average mortality.

$$R_k = \frac{\sum_x \sum_y \sum_i \sum_j D_{ijk}^{xy}}{\sum_x \sum_y \sum_i \sum_j T_{ijk}^{xy} n_j} \Big/ M \qquad [3.3]$$

The sum over x, y, j, and k (Equ. 3.4) gives the increase in mortality by months since last birthday R_i,

$$R_i = \frac{\sum_x \sum_y \sum_j \sum_k D_{ijk}^{xy}}{\sum_x \sum_y \sum_j \sum_k T_{ijk}^{xy} n_j} \Bigg/ M \qquad [3.4]$$

while the sum over x, y, i, and k (Equ. 3.5) results in *Rj*, the ratio of the deaths rates by current month as compared to average mortality.

$$R_j = \frac{\sum_x \sum_y \sum_i \sum_k D_{ijk}^{xy}}{\sum_x \sum_y \sum_i \sum_k T_{ijk}^{xy} n_j} \Bigg/ M \qquad [3.5]$$

Equ. 3.6 gives the age-specific death rates for those born in June (k=6) and December (k=12)

$$Q_{jk=6/k=12}^{x} = \frac{\sum_x \sum_y D_{ijk}^{xy}}{\sum_x \sum_y T_{ijk}^{xy} n_j} \qquad [3.6]$$

and the sum over all age groups x and time y results in the mortality trajectories over the time span of one year of life depending on the month of birth

$$R_{ijk} = \frac{\sum_x \sum_y D_{ijk}^{xy}}{\sum_x \sum_y T_{ijk}^{xy} n_j} \Bigg/ M \qquad [3.7]$$

The results of Equations 3.3 to 3.7 are displayed in Figures 3.1.A to 3.1.C. As expected, mortality increases by about 10% for males and females combined within twelve months (Fig. 3.1.A, solid lines females, dashed lines males). Mortality is lowest in August and highest in January; the difference is about 17.6% for males and 13.9% for females. The maximum difference in the death rates by month of birth is approximately 5.3% for males and 4.5% for females. The monthly pattern is similar for females

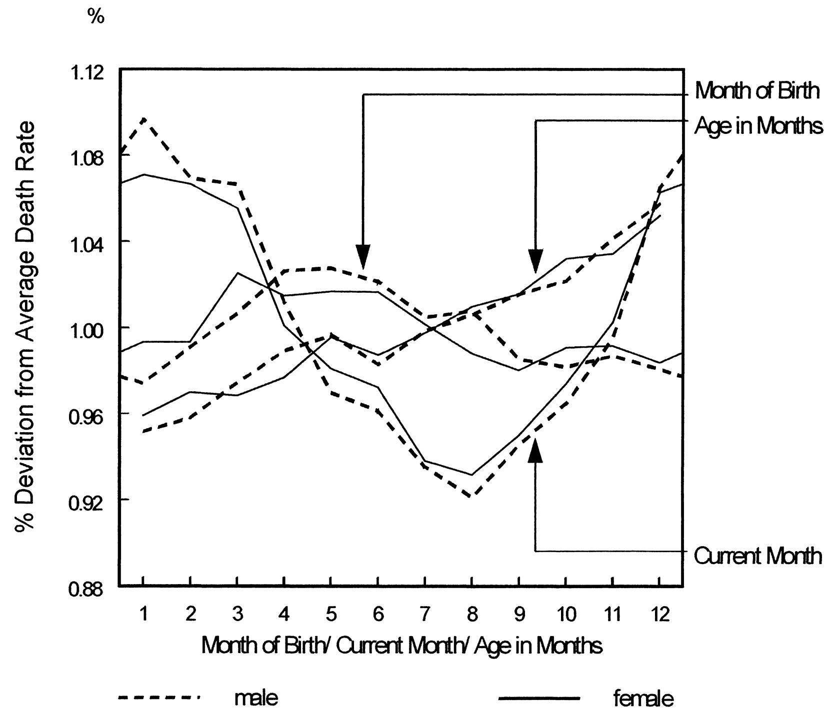

Figure 3.1.A Danish register data: deviations in standardized death rates by month of birth, current month and age in months.

and males. All three factors examined influence adult mortality simultaneously. This can be seen in Fig. 3.1.B and 3.1.C, where the mortality trajectories depend on month of birth, current month, and age in months. In Figure 3.1.B mortality differs considerably, depending on whether a person is born in June or December and therefore reaches a certain age during winter, when mortality is high, or during summer, when mortality is low. Figure 3.1.C shows the mortality trajectories over 12 months by month of birth. All three effects exist simultaneously and may interact with each other. In the following, an event-history model is used to estimate the force of mortality dependent on month of birth and corrected for interaction between month of birth and current half-year. The outcome of a model without the interaction term is described in Doblhammer (1999).

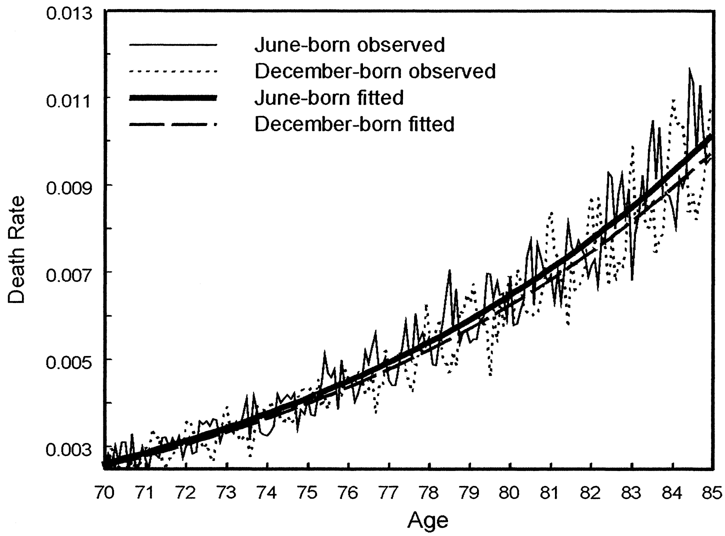

Figure 3.1.B Danish register data: death rates by age for June and December-born.

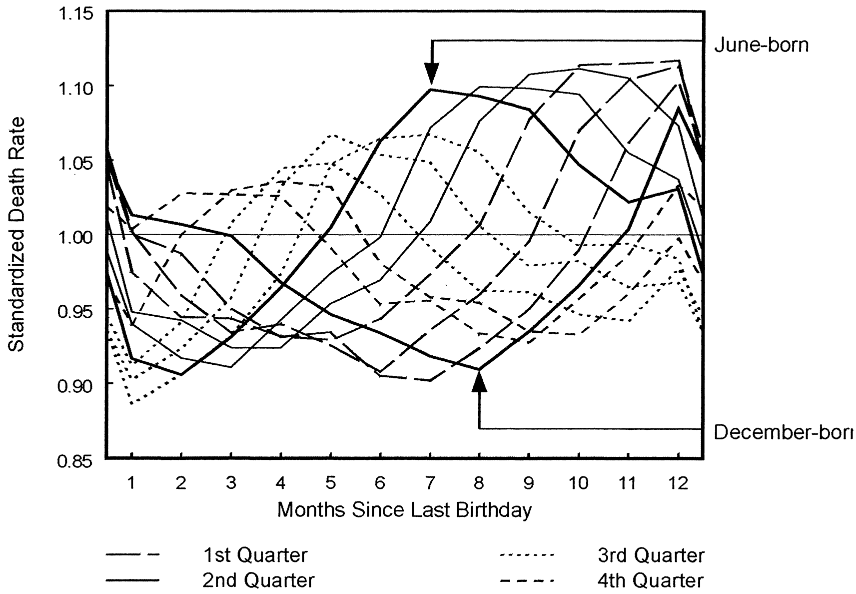

Figure 3.1.C Danish register data: death rates over the course of one year depending on the month of birth.

3.1.2 An Event-History Model to Separate the Effect of Month of Birth, Age, and Current Month

The longitudinal nature of the Danish dataset allows us to model $\mu(x)$, the force of mortality at age x, directly. The general mathematical specification of the model is

$$\mu(x) = \mu_0(x)\exp\left\{\sum_i \beta_i y_i(x)\right\} \qquad [3.8]$$

where $\mu(x)$ is a function of the baseline hazard $\mu_0(x)$; $y_i(x)$ denotes the value of the covariate and β_i the regression coefficient that measures the effect of the covariate *i* on the force of mortality. In this model the baseline hazard $\mu_0(x)$ is piecewise constant. The model assumes that the age-specific death rate for five-year age groups j follows a step function and that the death rate within the five-year age groups is constant.
Thus,

$$\mu_0(x) = a_j$$

$$\text{with} \qquad [3.9]$$

$$\text{j} = \text{50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-95,95+}$$

All other covariates are sets of binary variables that represent the levels of categorical covariates. The following covariates are included: sex; month of birth, which divides the year into thirteen periods, each consisting of four weeks. Week one starts at day one, week two at day eight, and so forth; current half-year, which is measured in two 6-months periods from January to June and July to December; and birth cohort: 1860-1879, 1880-1889, 1890-1899, 1900-1909, 1910-1918.

Month of birth and birth cohort are independent of time. The variable current half-year is time-varying and constructed such that in each half-year 50% of the average annual deaths occur. In the first half-year those with birth dates from January to June will be the oldest and will therefore experience increased mortality. In the second half-year the July to December-born will be the oldest and their mortality will be increased. Since in the first and second half-year the mortality risk is approximately the same the month of birth effect should disappear when an interaction effect be-

Table 3.1. Model 1: Relative mortality risks estimated by the hazard model defined in Equ. 3.8.

Variable	RMR	p-value	Variable	RMR	p-value
Month of birth		<.001	*Sex*		<.001
1.1–28.2.	1(RG)		Males	1(RG)	
29.1.–25.2.	1.007		Females	0.664	
26.2.–25.3.	1.031		*Cohort*		<.001
26.3.–22.4.	1.033		1860-1879	1(RG)	
23.4.–20.5.	1.034		1880-1889	0.969	
21.5.–17.6.	1.034		1890-1899	0.896	
18.6.–15.7.	1.026		1900-1909	0.842	
16.7.–12.8.	1.012		1910-1918	0.782	
13.8.–9.9.	1.009				
10.9.–7.10.	1.003		*Age*		<.001
8.10.–4.11.	0.999		50-54	1(RG)	
5.11.–2.12.	0.997		55-59	1.533	
3.12.–31.12.	0.996		60-64	2.373	
Current half-year			65-69	3.641	
1st half-year	1 (RG)		70-74	5.648	
2nd half-year	0.995		75-79	8.767	
			80-84	13.615	
			85-89	21.521	
			90-94	33.927	
			95+	52.818	

RG... reference group, RMR... relative mortality risk

tween month-of-birth and current half-year is constructed. The survival models were estimated with the program Rocanova.

Model 1 (Table 3.1) shows the main effects of the included variables without the interaction between month of birth and current month. The main effect of the month of birth reveals a 3 per cent excess in mortality of the March to July-born as compared to the January-born. Mortality decreases continuously over cohorts and is 22 per cent lower in the youngest cohort (1910-1918) than in the oldest cohort (1860-1879). The mortality of females is 64% of the mortality of males, the first and second half-year have equal mortality risks. The age-specific force of mortality approximately doubles with every 10-year age group.

Table 3.2. Model 2: Relative mortality risks estimated by the hazard model defined in Equ. 3.8. with an interaction term between month of birth and current half-year.

Variable	RMR	p-value	RMR	p-value
	Month of birth x current 1^{st} half-year		Month of birth x current 2^{nd} half-year	
Month of birth		<.001		<.001
1.1.–28.2.	1(RG)		1 (RG)	
29.1.–25.2.	1.025		0.993	
26.2.–25.3.	1.103		0.978	
26.3.–22.4.	1.149		0.952	
23.4.–20.5.	1.147		0.956	
21.5.–17.6.	1.120		0.974	
18.6.–15.7.	1.062		1.000	
16.7.–12.8.	1.012		0.011	
13.8.–9.9.	0.970		0.037	
10.9.–7.10.	0.962		0.031	
8.10.–4.11.	0.974		1.016	
5.11.–2.12.	0.965		1.019	
3.12.–31.12.	0.970		1.014	
Sex		<.001		
Males	1(RG)			
Females	0.644			
Cohort		<.001		
1860-1879	1(RG)			
1880-1889	0.968			
1890-1899	0.895			
1900-1909	0.842			
1910-1918	0.781			
Age		<.001		
50-54	1(RG)			
55-59	1.532			
60-64	2.370			
65-69	3.636			
70-74	5.640			
75-79	8.755			
80-84	13.598			
85-89	21.495			
90-94	33.890			
95+	52.765			

RG... reference group, RMR... relative mortality risk

Model 2 (Table 3.2) includes an interaction term between month of birth and current half-year. Taking again the January-born as the reference group one finds that in the first current-half year the February to August-born experience higher mortality and the September to December-born lower mortality. In the second half-year the picture is reversed. The August to December-born experience higher mortality while the January to July-born experience lower mortality. This is the expected pattern since in the first half-year those with birth dates January to July are the oldest, while in the second half-year those born between August and December are the oldest. The important result is that in the first half-year the excess mortality of those born in the first-half year is much higher (April and May-born: plus 15 % as compared to the January-born) than the excess mortality in the second half-year of those born in the second half-year (September and October-born: plus 4 % as compared to the January-born). Since approximately 50% of the deaths occur in each half-year the larger excess mortality of the spring-born cannot be the result of seasonal changes in the mortality risk but is rather caused by the month of birth.

In the following a third approach is presented that is based on death rates. The specification of the model varies from above hazard model. While in the hazard model an interaction term between month of birth and current half-year is used to explore the impact of the seasonal changing mortality risk on the month of birth effect, an interaction term between age since last birthday and current month is used in the next model.

3.1.3 A Model Based on Death Rates to Separate the Effect of Month of Birth, Age, and Current Month

For the Danish population it is possible to construct a surface of death rates by month of birth, current month and age in months using Equ. 3.1 and 3.2. The surface of the logarithm of the monthly deviations from the annual death rates is then modeled by the following Equ.:

$$\ln\left(R_{ij}^{xy}\right) = a_0 + \sum_{k=1}^{12} a_k I_{1k} + \sum_{k=1}^{12} b_k I_{2k} + \sum_{k=0}^{1} \alpha_k I_3 + \beta A + + \sum_{k=1}^{12} \gamma_k A\, I_{2k} + u \qquad [3.10]$$

Table 3.3. The logarithm of the relative deviations from yearly Danish death rates by month of birth, age in months, current month, and sex.

Covariates	Parameter estimates	Covariates	Parameter estimates
Month of birth		*Current Month*	
January (RG)	0	January (RG)	0
February	0.008 *	February	-0.013
March	0.027 **	March	-0.022 **
April	0.032 **	April	-0.066 **
May	0.033 **	May	-0.098 **
June	0.035 **	June	-0.106 **
July	0.016 **	July	-0.131 **
August	0.014 **	August	-0.145 **
September	0.004 **	September	-0.126 **
October	0.003 **	October	-0.106 **
November	0.008 **	November	-0.079
December	-0.003	December	-0.021
Sex			
Male (RG)	0		
Female	-0.025 **		
Age	0.075 **		
Constant	-.047		

** $p < 0.01$, * $p < 0.05$, RG denotes the reference groups.

In Equ. 3.10 the variables I_1 to I_3 are indicator variables. They take the value one for a particular characteristic and zero otherwise. I_{1k} indicates month of birth (reference month: January), I_{2k} current month (reference month: January), I_3 sex (reference sex: males). The variable A stands for age since last birthday in months rounded up to the nearest integer. Two models are estimated. The first model only includes the main effects of the variables and results are displayed in Table 3.3. The second model includes the interaction term $\sum_{k=1}^{12} \gamma_k A\, I_{2k}$, which estimates the increase in the seasonal mortality risk with age (Table 3.4). The parameter values a_1, b_1, α_0 and γ_1 are set equal to zero; $u \sim N(0, \sigma^2)$.

Table 3.4. The logarithm of the relative deviations from yearly Danish death rates by month of birth, age in months, current month, sex and an interaction term between age in months and current month.

Covariates	Parameter estimates	Covariates	Parameter estimates
Month of birth		*Current month*	
January (RG)	0	January (RG)	0
February	0.009 **	February	0.006
March	0.027 **	March	-0.009 **
April	0.033 **	April	-0.043 **
May	0.033 **	May	-0.084 **
June	0.036 **	June	-0.089 **
July	0.018 **	July	-0.113 **
August	0.015 **	August	-0.130 **
September	0.005 **	September	-0.108 **
October	0.006 **	October	-0.081 **
November	0.011 **	November	-0.060 **
December	-0.001	December	-0.009 **
Sex		*Age*current month*	
Male (RG)	0	January (RG)	
Female	-0.025 **	February	-.0029
Age	0.010 **	March	-.0019
Constant	-.0645	April	-.0036
		May	-.0022
		June	-.0026
		July	-.0027
		August	-.0023
		September	-.0027
		October	-.0040
		November	-.0028
		December	-.0017

** $p < 0.01$, * $p < 0.05$, RG denotes the reference groups

As compared to previous two analyses of month of birth, age, and month of death (current month) the results only change slightly when the impact of the different factors is estimated based on the model described in Equation 3.3. Those born between April and June experience excess mortality of about 3% as compared to the January-born. Current mortality is about 15% lower in August than in January (Table 3.3). Introducing the interaction effect between current month and age does not change the parameter estimates for the month-of-birth effect (Table 3.4) and mortality of the April to June-born is still 3 % higher than of the January-born. This result again confirms that the differences in life span by month of birth are not caused by the seasonal fluctuations in mortality.

3.2 The Procreational Habits Hypothesis

The second hypothesis assumes that the causal mechanism is linked to socio-economic differences in the seasonal distribution of births. With only a few exceptions births are distributed seasonally over the year. It has been shown that both social and biological factors influence this seasonality (Lam & Miron 1996, Doblhammer et al. 2001). The hypothesis is that, if the seasonality in birth is partly driven by the preference of couples for certain seasons, then the intensity of the preference may differ between social groups.

In schizophrenia research this hypothesis is known as the procreational habits hypothesis. In their review Torrey at al. (1997) mention Huntington (1938) as the first to explain the seasonal birth pattern in individuals with schizophrenia by the idiosyncratic seasonal conception pattern of their parents. And they cite James (1978), who describes the exceptional conception pattern as

"in the summer, people wear fewer clothes in bed, and... a schizoid spouse is more likely then to notice his (or her) co-spouse there and accordingly to initiate sexual behaviour".

Other contemporary proponents of this theory are Hafner et al. (1987) and Bleuler (1991), who proposes that conceptions among depressed individuals are more likely to take place in spring, when depression is less severe. In their review Torrey et al. come to the conclusion that the procreational habits theory can be ruled out as an explanation for the winter-spring excess births among schizophrenics. First, the theory predicts that individuals with schizophrenia who have a family history of the disease should have a greater winter-spring birth excess, which is clearly not the case. Second, the procreational habits theory also predicts that siblings of individuals with schizophrenia should also have a winter-spring birth excess. Torrey et al. (1997) cite six studies that did not find a winter-spring birth excess among siblings of people with schizophrenia.

To test this hypothesis in the context of mortality, education was used as an indicator of social group. Parental education is not contained on birth certificates from the beginning of the 20th century. It was thus assumed that, in 1991, a person's educational level was closely linked to the educational level of his or her parents. The Austrian microcensus of the year 1991 reveals that the educational status of the 15- to 19-year-old Austrians depends to a large extent on the social and educational status of the parents (Schwarz et al. 2002a, Schwarz et al. 2002b), despite tuition-free access to all levels of education since the 1970s. For earlier birth cohorts that did not benefit from the expansion of the Austrian education system, the intergenerational correlation in education must have been even stronger than it is today.

We calculated the seasonal distribution of the birth dates of Austrians aged 50+ by educational group on the basis of the 1981 census. It appears that the spring peak in births is stronger among adults with high or medium education, while people with basic education are over-represented among the autumn-born (Fig. 3.2).

The correlation between mean age at death by month of birth and the deviations in the monthly birth distribution from the average monthly pattern is -0.78 (Pearson correlation, one-sided test: $p<0.001$) for adults with high education, -0.84 ($p<=0.001$) for adults with medium education, and 0.88 ($p<0.001$) for adults with basic education. A recent study on the seasonality of births in the former Czechoslovakia comes to the same results (Bobak & Gjonca 2001). A similar result was found on the basis of a 10% sample of the 1971 census of economically active, British-born males. Non-manual workers tended to be born in spring and manual workers in autumn and winter (Smithers et al. 1984).

One can therefore conclude that the month-of-birth pattern in longevity is not due to differences in the conception pattern by socio-economic group. On the contrary, it appears that the differences in life span would be even larger if socio-economic differences were taken into account.

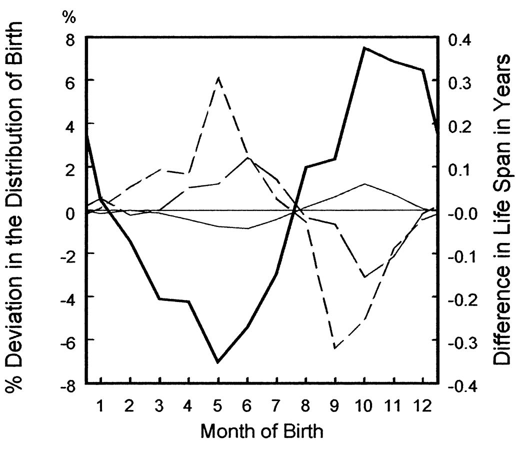

Figure 3.2. Per cent deviations in the seasonal distribution of births of Austrians (ages 50+) by education from the average seasonal distribution of births: medium (small dashed line), high (long dashed line), basic education (solid line) and differences in mean age at death (bold solid line).

3.3 The Deadline Hypothesis

Another prominent hypothesis in sociological and psychological research is that children who are born after a certain deadline have to wait another year before they can enroll at school, which makes them about a year older than the youngest children in their school class. For example, at the beginning of the 20th century school started on 1 October in Austria. Those children who had not turned six before this date had to wait another year. Thus, children born in the autumn experienced an age advantage over their classmates.

Research suggests that this age advantage affects scholastic aptitude, which might translate into a lifelong advantage in various regards. Alton and Massey (1998) studied the General Certificate of Secondary Education (GCSE) examinations. Almost all pupils in England, Wales, and Northern Ireland take this examination at age 16. Starting with pupils born in September and moving to those born in August, i.e. proceeding from the oldest to the youngest pupils of a given school year, the authors find that there is a steady decline in marks. By the age of 18 these differences have vanished. The authors attribute this to a selection effect stemming from the fact that only those students who did well at the GCSE examinations will have continued to the advanced level. A significant interaction between reading failure and age category (overage at school entrance vs. correct age) by season of birth was observed amongst second-year pupils (Flynn et al. 1996). Although the authors mention entrance cutoff birth dates to kindergarten as a probably explanation they do not rule out that the reported associations may be attributed to selective sampling.

The effect of relative age has been widely tested in the area of sports. Dudink (1994) finds a significant relationship between date of birth and success in tennis and soccer. In both disciplines half of the tennis players in the top rankings of the Dutch youth league and a third of the English soccer players in league clubs were born during the first three months of the competition year. The consistent asymmetry in the birth-distribution of senior professional soccer players has led Helsen et al. (1998) to investigate whether such asymmetries already exist in youth categories in soccer. Their results indicate that young players born from August to October are more likely to be identified as talented at ages 6 to 8 and are exposed to higher levels of coaching. These players are born early in the selection year and are relatively older. Those born relatively late in the selection year tend to dropout as early as 12. A study of the effect of a horse's month of birth on its future performance in sport (Langlois & Bloin 1998) showed that, under the age of four, horses that were born in the second part of the year were particularly disadvantaged in flat races. The disadvantage was not so strong for jumping races and for trotters. After the age of four the effect was of much smaller magnitude. They suggest that two effects may be responsible for the results: a relative age-effect, which is more prominent in races where older horses are favoured. This relative-age effect decreases with age (probably due to selection). The second effect is seen mainly among older horses in equestrian competitions and is not further specified by the authors.

Angrist and Krueger (1991) performed the most widely known study exploring the deadline effect with regard to outcomes in schooling and earnings. However, the authors argue the other way round from the studies cited above. Since compulsory schooling laws require students to stay in school until their 16th or 17th birthday, the oldest students may drop out slightly earlier, which should be reflected in their later earnings. In their study they exploit an "unusual natural experiment" to estimate the effect of compulsory schooling laws on educational attainment and earnings. They explain the experiment as follows:

"The experiment stems from the fact that children born in different months of the year start school at different ages, while compulsory schooling laws generally require them to remain in school until their sixteenth or seventeenth birthday. In effect, the interaction of school-entry requirements and compulsory schooling laws compel students born in certain months to attend school longer than students born in other months. Because one's birthday is unlikely to be correlated with personal attributes other than age at school entry, season of birth generates exogenous variation in education that can be used to estimate the impact of compulsory schooling on education and earnings" (p. 980).

In their study they find that children born in the first quarter of the year have a slightly lower average level of education than children born later in the year. According to the authors, school districts typically require a pupil to have turned six by January 1 of the year in which he or she enters school. Therefore, those born at the beginning of the year enter school at an older age and attain the legal dropout age earlier in their educational careers than people born later in the year. In support of their study they report that the seasonal pattern of education is not evident in college graduation rates or in post-graduate degree completion rates. Since compulsory schooling laws do not affect these two types of institutions, the authors conclude that the relationship between years of schooling and month of birth is entirely due to compulsory schooling laws. In the second part of their paper they showed that not only educational attainment but also earnings later in life are related to the month of birth.

In a recent paper by Pflug (2001) the author repeats the study by Angrist and Krueger for the Netherlands and presents evidence that the season of birth influences school performance not in terms of compulsory schooling effects but in the area of learning abilities. They refer to the arguments presented above – that it is the relative age among class mates that influences learning ability and, as a result of education, earnings later in life.

Further criticism of Angrist and Krueger's study was raised by Bound and Jaeger (1996). They showed that the association between quarter of birth and earnings or other labour-market outcomes existed for cohorts that were not bound by compulsory school attendance laws. Furthermore, Angrist and Krueger did not take into account the fact that differences in the birth distribution of social groups exist and that other factors, such as health, may be related to the season of birth.

Based on the studies discussed above, there seems to exist firm evidence that the relative age among pupils influences their scholastic aptitude and their sporting success. In Austria at the beginning of the 20th century school started on 1 October and children who had not turned six before this date had to wait another year. Thus, the autumn-born were the oldest among their classmates. Nevertheless one finds that the autumn-born achieve lower educational levels than the spring-born, which is contrary to the evidence presented above. In addition, at the turn of the 20th century in Austria, many schools, particularly in rural areas, combined several years into one class and a single teacher taught the different levels.

3.4 Selective Survival or Debilitation

The fourth hypothesis assumes that selective survival during infancy is the causal mechanism that explains the relationship between month of birth and life expectancy. Specifically, the hypothesis is that autumn-born infants suffer higher mortality in their first year of life than spring-born infants. This would leave the relatively more robust individuals alive, who would experience lower mortality at adult ages.

Infant mortality decreases dramatically over the first year of life. Looking at births and infant deaths in Denmark between 1911 and 1915, the probability of death decreases from 0.023 in the first 24 hours of life to 0.00275 in the twelfth month of life. At the same time, infant mortality is highly seasonal. However, the pattern of seasonality depends on age. Although at almost all ages death rates are highest in the winter months in the

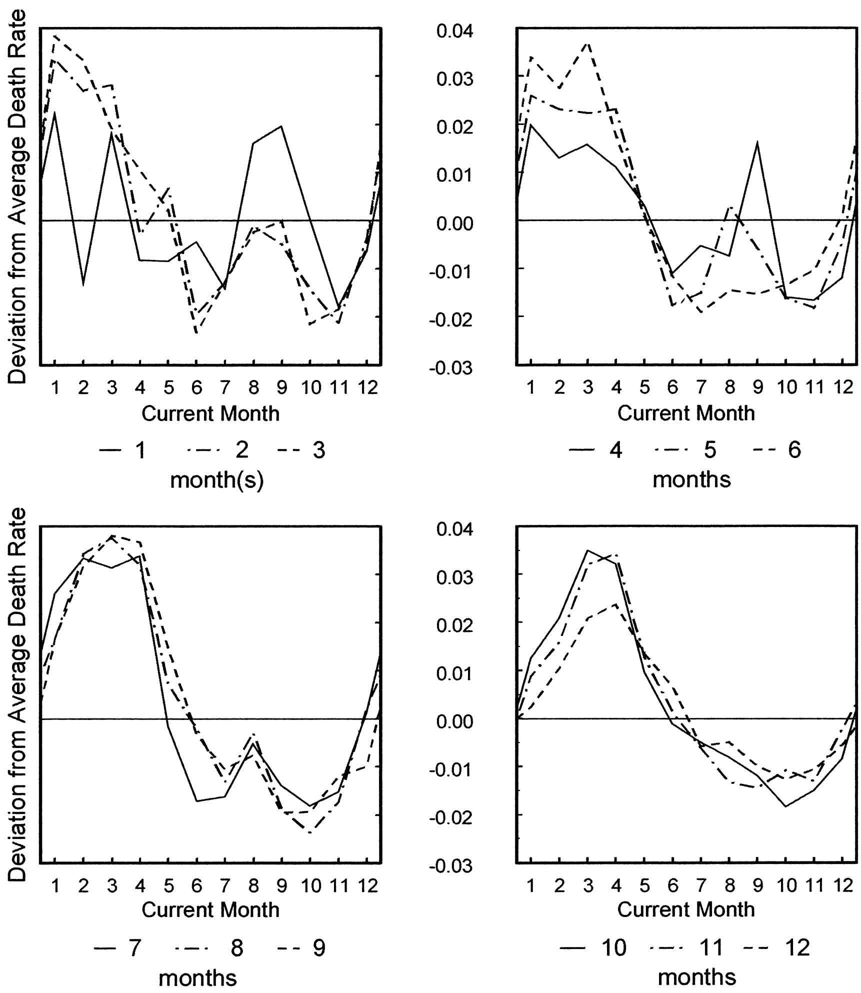

Figure 3.3. Deviation from the average death rate at a particular age (in months) by month of the year (current month) in Denmark 1911–1915.

first five months of life, a prominent secondary peak exists for the summer months of July, August, and September. This secondary peak tends to become smaller and even to disappear from the sixth month of life onwards (Fig. 3.3). In a model of the effect of the month of birth on survival during the first year of life one therefore has to account for the changes in the seasonality of deaths with age and for the strong decrease in the death rate over the first twelve months.

To test whether selective survival or debilitation during the first year of life explains the differences in remaining life expectancy at age 50, data on infant mortality for Denmark aggregated over the five-year period 1911-1915 were used. The tabulations in the "Statistik tabelvaerk " provide the number of deaths by age of the infant in months and month of death. Thus, the actual date of birth of the deceased infants can fall within a period of two months – on average it falls on the first day of each month. On the basis of the monthly number of births aggregated over the period 1911–1915, death rates for each month of the first year of life were then calculated. In order to center the number of births at the first day of each month the average number of births of two adjacent months are used in the calculations.

To estimate the impact of month of birth on the monthly death rates in the first year of life a model was used that is similar to that employed for adult ages. The logarithm of the monthly deviations from the annual death rate was modeled by applying equation [3.11]:

$$\ln\left(R_{ij}\right) = a_0 + \sum_{k=1}^{12} a_k I_{1k} + \sum_{k=1}^{12} b_k I_{2k} + c_1 \ln x + \sum_{k=1}^{12} \alpha_k \ln x I_{2k} + u \qquad [3.11]$$

The variables I_1 and I_2 are indicator variables. They take the value one for a particular characteristic and zero otherwise. I_{1k} indicates month of birth (reference month: January), I_{2k} current month (reference month: January). Age x is in months rounded up to the nearest integer. To reduce the number of parameters in the model and because the decline in mortality during the first year of life can be modeled by a logarithmic function, x is included as a numerical variable. The term $\sum_{k=1}^{12} \alpha_k \ln x \, I_{2k}$ corrects for the interaction in the death rates between age and current month. The infant mortality surface consists of only 144 data points (12 birth months x 12 ages). Monthly births are only available for both sexes combined and infant deaths by age in months and month of death are aggregated over the five-year period 1911-1915. Hence, the model does not include sex and cohort effects. The parameter values a_0, a_k, b_k, c_k and α_k were estimated by least squares regression; u~N(0,σ^2).

When applied to Danish data on seasonal infant mortality in the first year of life from the years 1911 to 1915, the model in Equ. 3.11 seems to indicate that infants born in June are the most vulnerable. Table 3.5 contains the parameter estimates of the model described in Equ. 3.11 and those of a model without correction for the interaction between month of death and age. Both models fit the data well, with an adjusted R^2 of 0.93 for the model without correction of age at death and current month and an adjusted R^2 of 0.96 for the model with correction. The latter model predicts that the standardised death rate during the first year of life is about 31% higher for the June-born than for the February-born. The correction for the interaction between the seasonality of death and age has two effects. First, it decreases the mortality disadvantage of the June-born and, second, it shifts the whole pattern to the left.

This is not the first study to recognize that infant mortality depended strongly on the season of birth historically and that the seasonal pattern of infant deaths depends on age. Already in 1938 Huntington presented similar results for Belgium and New York (Huntington 1938). Breschi & Bacci

Table 3.5. Effect of the seasonality of death on infant mortality by month of birth.

	Model 1		Model 2	
	Parameter estimate	p-value	Parameter estimate	p-value
Intercept	-1.529	0.0000	-1.352	0.0000
Current month				
January	0.358	0.0000	0.241	0.1462
February	0.421	0.0000	0.084	0.6175
March	0.506	0.0000	0.050	0.7673
April	0.451	0.0000	-0.165	0.3345
May	0.217	0.0052	-0.147	0.3916
June	-0.017	0.8188	-0.339	0.0498
July	-0.122	0.1122	-0.241	0.1616
August	-0.052	0.4955	-0.068	0.6898
September	-0.174	0.0241	0.072	0.6715
October	-0.291	0.0002	-0.033	0.8433
November	-0.232	0.0028	-0.060	0.7164
December	0.000		0.000	
Month of birth				
January	-0.072	0.3488	-0.044	0.5058
February	-0.070	0.3572	-0.023	0.7308
March	-0.023	0.7672	0.018	0.7964
April	0.080	0.2983	0.095	0.1662
May	0.206	0.0078	0.165	0.0179
June	0.304	0.0001	0.233	0.0010
July	0.301	0.0001	0.195	0.0054
August	0.254	0.0011	0.136	0.0486
September	0.203	0.0088	0.081	0.2334
October	0.112	0.1436	0.025	0.7099
November	0.020	0.7904	-0.023	0.7328
December	0.000		0.000	
ln (age)	-0.854	0.0000	-0.938	0.0000

Table 3.5. (continued)

	Model 1		Model 2	
	Parameter estimate	p-value	Parameter estimate	p-value
*Interaction ln(age)*current month*				
January* ln(age)			0.071	0.4390
February*ln(age)			0.203	0.0305
March*ln(age)			0.274	0.0042
April*ln(age)			0.369	0.0002
May* ln(age)			0.218	0.0234
June * ln(age)			0.193	0.0448
July * ln(age)			0.071	0.4549
August * ln(age)			0.010	0.9197
September * ln(age)			-0.148	0.1179
October * ln(age)			-0.155	0.0970
November * ln(age)			-0.103	0.2574
December * ln(age)			0.000	

(1996) came to a similar conclusion for Switzerland and Belgium, as did Lumaa et al. (1998) for Finland and Reher & Sanz-Gimeno for Spain (2003).

If selective survival during infancy explains the differences in adult life span, then one would expect a significant negative correlation between the probability of death in the first year of life and after age 50. If debilitation *in-utero* or during the first year of life (our fifth hypothesis) is the causal mechanism, then one would expect a significant positive correlation because rates of debilitation are highly correlated with mortality.

For Denmark the correlation between infant mortality in the first year of life and adult mortality after age 50 (Figure 3.4) is 0.87 and statistically significant ($p<0.0001$). This result clearly rules out the selection-hypothesis and indicates that the causal mechanism of the month-of-birth pattern in life span is related to debilitating factors that affect either the mother during pregnancy or the infant in the first year of life.

3.5 Conclusion

There are four main conclusions that can be drawn from the analysis of data from Austria, Denmark, Australia, and Hawaii:

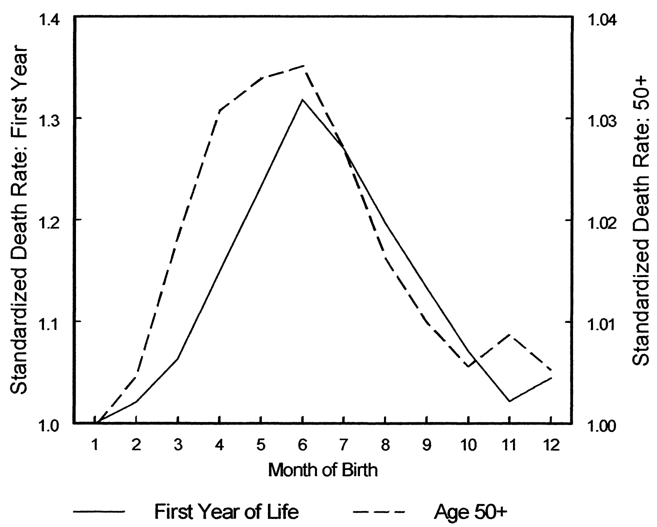

Figure 3.4. Standardized death rates of Danish infants born in a specific month and difference in remaining life expectancy at age 50 by month of birth.

First, the differences in life span by month of birth are tied to the seasons of the year. This is clearly shown by the reversal of the pattern in the Northern and the Southern Hemispheres.

Second, the pattern also exists in regions close to the equator: Hawaii is one example, Queensland in Australia is another. This finding rules out the explanation that differences in the length of daylight are the underlying causal mechanism. This hypothesis will be further explored on the basis of the US death data (Chapter 5).

Third, although Queensland and Hawaii are close to the equator these two countries experience two distinct seasons of the year, namely a wet and a dry season. In both countries those who experience a mortality advantage later in life are born at the beginning of the season when temperatures are more moderate. In non-tropical regions with four distinct seasons those born in autumn and winter have longer lives.

Fourth, within a country the pattern may vary greatly among different groups. This is shown by the analysis of the month-of-birth pattern for Aborigines in Australia. Not only is the peak-to-trough difference for them more than nine times as large as for other native-born Australians – the pattern is also different. The next chapter will explore whether this is also true for different social and racial groups in the United States.

The test of the five hypotheses clearly reveals that the month-of-birth pattern is not the result of an age-artifact caused by the seasonality of death, nor is it the result of social factors such as social differences in the seasonal distribution of births or age at entry into school. There is no support for the hypothesis that differential infant survival is the causal factor behind the observed phenomenon. On the contrary, for Denmark there exists a strong and highly significant positive correlation between the differences in infant mortality and the differences in mortality after age 50 by month of birth, which is consistent with the hypothesis that debilitation early in life is the causal mechanism.

The results support the theory of there being a critical period early in life that affects adult health and survival without other mediating life-course factors. The model of infant mortality does not permit any conclusions about whether debilitation occurred *in-utero* or during early infancy. The results are consistent with both the effect of seasonal differences in nutrition and of seasonal differences in the disease environment early in life. In past decades the food supply in general, and the availability of fresh fruit and vegetables in particular, differed from season to season. Mothers who gave birth in autumn and early winter had access to plentiful food and fresh fruit and vegetables throughout the third trimester of pregnancy, the period of peak growth of the fetus; those who gave birth in spring and early summer experienced longer periods of inadequate nutrition during their last trimester. On the other hand, infant mortality was historically primarily the result of infectious disease, which occurred seasonally. The close correlation between infant mortality by month of birth and the month-of-birth pattern in adult life span therefore suggests that infectious disease plays an important role.

In the Danish data, differences in adult life span by month of birth are significantly smaller in the more recent cohorts than in the oldest cohorts. In a previous study by Doblhammer & Vaupel (2000) it is shown that among those born before 1888 the differences are amplified by a factor of 72%. That is, the difference in adult mortality by month of birth is almost twice as large for the cohorts born from 1863 to 1888 as it is for those born between 1889 and 1918. Chapter 6 in this book provides a more detailed analysis of the cohort changes in the month-of-birth pattern. This shift lends further support to the explanation that the relationship between month of birth and adult life expectancy is caused by prenatal and early postnatal conditions related to nutrition and infectious disease, since considerable improvements in maternal and infant health took place between these two periods (Preston & Haines 1991).