The impact of World War II on the cancer rates in Norway

by

Anahita Rahimi

THESIS
for the degree of
MASTER OF SCIENCE

(Master i Modellering og dataanalyse)

Faculty of Mathematics and Natural Sciences
University of Oslo

February 2011

Det matematisk- naturvitenskapelige fakultet
Universitetet i Oslo

Dedicated to my beloved parents Mahboubeh Eftekhari and Reza Rahimi
and brother Mazdak Rahimi.

Preface

After a meeting with Bjørn Møller and Freddie Bray from the Cancer Registry of Norway and Ørnulf Borgan from the University of Oslo in December 2008 we decided that this thesis should be written in collaboration with the Cancer Registry of Norway. At this particular meeting Møller and Bray had two suggestions for topics, whereas one of the topics where chosen for this thesis. In earlier studies it has been shown that there is a transient effect of World War II on rates of colorectal, breast and testicular cancer, most probably due to the change in dietary and physical habits during the occupation period. Thus Møller and Bray where curious to see if a wartime effect could be found for other cancer sites in Norway as well.

Due to the fact that the incidence rates for cancer in Norway - and other countries - are growing I thought it might be interesting to see if the dietary and physical habits could significantly influence the incidence rates in certain epochs of time. This is also of concern to general public health. Thus "*The impact of World War II on the cancer rates in Norway*" was chosen as topic.

The Cancer Registry of Norway provided the data for the thesis. In addition the registry provided me with office space and supervising via Freddie Bray.

Acknowledgements

I would like to express my appreciation towards those people who made the completion of this thesis possible. My greatest gratitude goes to my supervisor Professor Ørnulf Borgan from the University of Oslo. With his great intelligence, kindness and patience he has helped me throughout the entire process of writing this thesis. Ørnulf, with his encouraging words and the amount of time spent on guidance, is an excellent supervisor. My gratitude also goes to my second supervisor Dr. Freddie Ian Bray, who with his encouragement and his excellent sense of humor has helped me to stay positive and enjoy the great amount of time spent on this thesis. In addition I am grateful for all the time and knowledge Freddie has shared with me. It has been a great honor working with both of you.

I would also like to show my appreciation towards my parents Mahboubeh Eftekhari and Reza Rahimi, with whom I have learned that with determination and hard work anything can be accomplished. Through their great amount of support and always believing in me, they have helped me to never give up, especially when things seem unbearable. Mom and Dad, thank you for everything.

To the rest of my family and wonderful friends, thank you so much for all your support and encouragement.

Last, but not least, I would like to thank the Cancer Registry of Norway, which gave me the opportunity of writing this thesis.

Oslo, February 2011
Anahita Rahimi

Contents

Chapter 1

Introduction

The second World War (WWII) involved most of the world's nations and lasted from 1939-1945. Several countries were occupied during this period. One of these countries was Norway, which was occupied during a five year period from 1940 to 1945. Due to the rationing of several food items during the occupation period, the dietary habits changed (e.g. Tretli and Gaard, 1996). While the intake of fresh vegetables, fish and potatoes increased in people's diet, the intake of energy, fat, meat and milk consumption decreased. As a result of the occupation period, tobacco and alcohol was not easily accessible, thus the consumption of these items was also reduced. In addition physical activity changed for the Norwegian population during the occupation period. Thus assumptions that changes in these factors might have affected the risk of cancer for selective cancer sites during the occupation period are present. Earlier studies have concluded with a transient reduction in incidence rates due to the impact of WWII for colorectal cancer, breast cancer for females and testicular cancer (Svensson et al., 2002; Tretli and Gaard, 1996; Wanderås et al., 1995). The decrease in risk for colorectal cancer was observed for birth cohorts born during and shortly after WWII. Similarly for breast cancer a decrease in the incidence rates was observed for the cohorts being in puberty during the occupation period. For testicular cancer, the decrease was observed for those born during the war, and it might seem that the cohorts being born just before the war also might have been affected (Wanderås et al., 1995). The three studies all imply that dietary habits are vital when it comes to risk of cancer, more specifically during early life for colorectal and testicular cancer and beginning of breast development at puberty and first full-time pregnancy for breast cancer for females. More specifically, Tretli and Gaard (1996) found a decrease for

women that were between eight and 27 years of age during the occupation period. In addition the study observed that the slope of the cancer rates for women being born between 1933-1944 had a tendency to level off after a strong increase.

Now if dietary habits *do* play a vital role in the risk for colorectal, breast and testicular cancer, a natural conjecture would be that it could play a vital role for other cancer sites as well. Thus we would like to investigate for other sites a possible decrease in cancer risk for birth cohorts born during WWII. In addition, we will consider birth cohorts experiencing puberty around WWII for females registered with breast cancer. These considerations are the motivation for the topic being addressed in this thesis, that is *the impact of World War II on the cancer rates in Norway.*

The Cancer Registry of Norway started recording cancer cases as early as 1952 (Cancer Registry of Norway, 2010a). It is mandatory to report all cancer cases to the Cancer Registry. Thus we trust the data being used in this thesis to be reliable and accurate. The data will be used for both visual inspections and statistical tests. It would be too time consuming and not of any purpose to try the methods on all the sites in question before being somewhat certain that the methods are reliable. Thus whenever examples need to be given to illustrate a methodology, data for colon cancer by sex will be used. When feasible, data for breast cancer for females and testicular cancer may be used as well. This is due to the fact that earlier studies have concluded with a transient reduction in the incidence rates for the birth cohorts around WWII for these specific cancer sites. Thus visual inspections and statistical tests should be able to capture this feature for these specific sites if we should trust them to give us reasonable results for the other cancer sites as well.

As mentioned above we hope that inspection of data will help us get a better overall view of the trends in incidence rates around WWII. The studies by Svensson et al. (2002), Tretli and Gaard (1996) and Wanderås et al. (1995) found that WWII has had an impact on the estimated incidence rates for the specific cancer sites considered. However these studies did not test for significant wartime effects. We will introduce such formal tests in this thesis and hope to verify a wartime effect beyond visual inspections. However we are aware of the possibility that the relatively small population size in Norway might be a drawback for the analysis part of this thesis.

Calculations and graphics in this thesis were obtained using **R** (R Development Core Team, 2010). One of the advantages of using **R** is that we easily can implement different packages in the software. The packages are developed to be used in the different fields of statistics and hence with func-

tions not given directly in the software itself. As for this thesis the so-called Epi package is ideal. The package contains functions which can be used for both visual inspections and the statistical tests considered in this thesis. We will not go into further details of the functions or other details regarding the software here. However, when needed, we will specify which functions we consider from the Epi package for the statistical tests and visual inspections considered in the following chapters.

The outline of the thesis is given as follows. In Chapter 2 we give an overall summary of the data. The chapter also gives details on how to define *birth cohort* by age and period and also gives a graphical presentation of this by introducing the *Lexis diagram*. A summary of the *cancer sites* considered in this thesis is also given in Chapter 2. In addition colon cancer is used as an example where the number of *new cases*, *person-years* and *incidence rates* given by sex are given in appropriate tables. Figures of observed rates given by period by age and birth cohort by age, also for colon cancer, are given for a better understanding of how to observe a period or cohort effect. Thus the purpose of Chapter 2 is mainly to give background information so the reader better will understand the methods and interpretations given in the following chapters.

In Chapter 3 we introduce the *age-period-cohort model* (apc model). The apc model is a Poisson regression model which considers age, period and cohort effects simultaneously. The three variables are hopelessly entangled since cohort is obtained by subtracting age from period. Due to the linear dependency between the three variables the use of the apc model and interpretations of the results should be handled with care. Necessary details for a better understanding of the model and its results are given, although the reader should consider for example Holford (1991) or Bray (2005) for further details. Visual inspections of the estimated effects from the apc model for colon cancer by sex are also considered in the chapter. The apc model can be used directly on 5-year age and period intervals. However, as the occupation period lasted for five years, we examine estimated effects by using yearly data as well as an aid to interpretation. Thus we introduce the term *splines*, which are integrated in the apc model for smooth estimated effects when using yearly data. Furthermore the model is the foundation of both the visual inspections and statistical tests introduced later in this thesis.

Chapter 4 introduces two tests which may help us give more formal conclusion in our interpretations of the estimated cohort effects for the cancer sites discussed in this thesis. Both tests were introduced by Tarone and Chu (1996, 2000). The first test can be seen as a generalization of *second differences*. Thus a recapitulation of the method of second differences is also

given in the chapter. Basically, the first test examines the non-linear cohort effects around WWII by considering two scenarios. In the first scenario, we assume linear slopes in two adjunct time intervals. In the second scenario we assume the estimated effects to be given as a curvature in a coherent time interval. In both scenarios we examine how the estimated cohort effects, given as linear slopes or curvature, change during the time around WWII. In both scenarios we hope to find a transient reduction in the estimated effects around WWII. We also compare numerical results, for colon, breast and testicular cancer, for both scenarios in this chapter. The second test is a nonparametric test which is a generalization of the sign test and is based on observed rates. However, the authors Tarone and Chu suggest that the test is used as a adjunct to the apc model introduced in the previous chapter.

In the fifth chapter, we present numerical results for all cancer sites considered in this thesis by using the first test introduced in the previous chapter. The results will be given in appropriate tables and figures. We hope that the results obtained in Chapter 5 will help us gain more strength in our conjecture of a wartime effect on the incidence rates for some cancer sites in Norway.

In the sixth and final chapter we will sum up the main findings in this thesis. We will give room for discussion and proposals for further research.

Chapter 2

Routine sources of data

In this chapter we give a summary of the data used in this thesis. The summary involves details of the variables available in the data extracted from the Cancer Registry of Norway. To better understand how birth cohorts are defined, the Lexis diagram will be introduced. The diagram graphically shows how a birth cohort is given by age and period. For illustration tables of the number of new cases and incidence rates are given for colon cancer. Visual inspections for the observed rates given by period by age and birth cohort by age for colon cancer, by sex, are also given.

2.1 Summary of data and sites

To make the analysis as good as possible we extract registered cases for 19 of the most common cancer sites given in Table 2.1. For each site the data contain the number of *new cases* and *person-years* for a given year, by age and sex. To better understand the definition of person-years we may consider 1000 individuals for a time period of 1 year (Scenario 1) and 500 individuals for a time period of 2 years (Scenario 2). For Scenario 1 we calculate the person-years by 1000 individuals $\times$ 1 year and similarly for Scenario 2 the calculation is given as 500 individuals $\times$ 2 years. Thus for both scenarios the person-years are equal to 1000. More formally we define person-years as the sum total of length of time a group of people are at risk for a given period, by age and sex. Data are available for both 1- and 5-year age and period intervals. From the yearly data we may easily obtain data with 2-year age and period intervals as well. The choice of dataset in the different settings of visual inspection and statistical tests will be specified when needed.

Regardless of the dataset the youngest and oldest age groups will have very few or zero observed number of new cases. To avoid irregularities and misinterpretations of the visual inspections we omit the youngest and oldest age groups. Thus we restrict the age interval for the cancer sites to be 30-69 years. For testicular cancer younger males are more at risk (Cancer Registry of Norway, 2010c) and the age interval will be restricted to 15-54 years for this particular cancer site. The age groups at risk for prostate cancer also deviate from the majority of cancer sites where there is almost zero incidence for those under the age of 40. For this site we restrict the age interval to 40-79 year.

Table 2.1: *The cancer sites considered in this study.*

ICD-10	**Site**
C00-14	Mouth and pharynx
C16	Stomach
C18	Colon
C19-21	Rectum, rectosigmoid and anus
C25	Pancreas
C33-34	Lung and trachea
C43	Melanoma of the skin
C50	Breast (for females)
C53	Cervix Uteri
C54	Corpus Uteri
C56	Ovary
C61	Prostate
C62	Testis
C64	Kidney excluding renal pelvis
C66-68	Bladder, ureter and urethra
C70-72	Central nervous system
C73	Thyroid gland
C82-85+C96	Non-Hodgkin lymphoma
C91-95	Leukaemia

2.2 Lexis diagram

The main purpose of this thesis is to study trends in incidence rates for *birth cohorts* around WWII. A graphical presentation of the relationship between age, period and cohort can be given by a Lexis diagram. The Lexis diagram

will be presented with 5-year age and period intervals. Interpretation and presentation of the diagram by using 1- and 2-year age and period intervals will basically be the same, except some minor adjustments to the length of the intervals and axis labels.

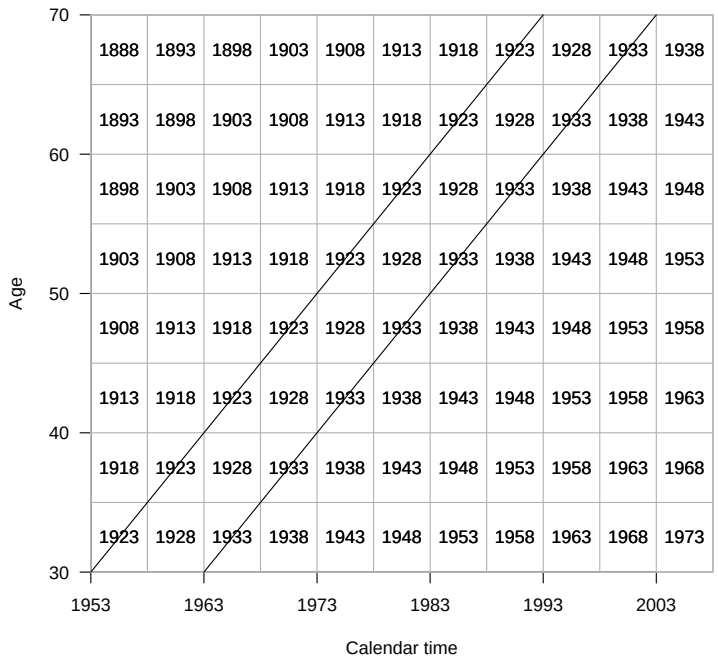


Figure 2.1: *Lexis diagram which shows the relationship between age, period and birth cohort using 5-year data. Period is given on the horizontal axis and age on the vertical axis. The birth cohorts can be seen on the diagonal, with a line going through the 1923 and the 1933 birth cohort, that is for those being born in 1918-27 and 1928-1937.*

Now for 5-year age and period intervals, the age groups considered are 30-34, 35-39, ..., 65-69 years and the periods are 1953-1957, 1958-1962, ..., 2003-2007. The respective birth cohorts are derived by subtracting age from period. As an example we subtract the oldest age group 65-69 from the first period interval 1953-1957. This leads to the cohort of people being born sometime in the 10-year interval 1883-1892. Hence the birth cohorts are given as the following 10-year overlapping intervals 1883-1892, 1888-1897, ..., 1968-1977. As a matter of notation we will denote the age groups and period intervals more briefly as 32.5, 37.5, ..., 67.5 and 1955.5, 1960.5, ..., 2005.5. That is, the mid-year will represent the 5-year interval of age and

period. The corresponding birth cohorts are also denoted by the midyear of the 10-year intervals, i.e. by 1888, 1893, ..., 1973.

In the Lexis diagram, see Figure 2.1, age is given on the vertical axis and period on the horizontal axis. Thus the respective birth cohort intervals are given following the diagonal up and towards the right. A line through the 1923 (1918 - 1927) and 1933 (1928 - 1937) birth cohorts are added so we can get at better feel of how the birth cohorts can be traced in the Lexis diagram. The Lexis diagram is easily made by the function *Lexis.diagram* in the Epi package in the software **R** (R Development Core Team, 2010).

As mentioned earlier we introduce the Lexis diagram so we can better understand the relationship between age, period and birth cohort. For further details, I will refer to the part about the Lexis diagram in Bray (2005).

2.3 Cancer rates

It will be helpful to examine incidence rates around WWII. 5-year age and period intervals for colon cancer, by sex, will be used for illustration. The tables and figures given in this section are constructed by the functions *stat.table* and *rateplot* in the Epi package in the software **R** (R Development Core Team, 2010).

We define the estimator for the incidence rate in age group a and period p by $\hat{r}_{ap} = \frac{d_{ap}}{Y_{ap}}$, where d_{ap} and Y_{ap} are the number of new cases and person-years for the corresponding age group and period. An overview of the number of new cases for both sexes are given in Table 2.2. Corresponding tables of person-years for both sexes are given in Table 2.3.

Table 2.2: *The number of new cases for colon cancer, by age and period.*

Male											
	Period										
Age	1955.5	1960.5	1965.5	1970.5	1975.5	1980.5	1985.5	1990.5	1995.5	2000.5	2005.5
32.5	8	13	12	6	12	15	25	15	12	15	23
37.5	17	26	17	18	15	29	22	30	17	29	29
42.5	36	33	39	32	29	38	51	62	69	58	74
47.5	44	53	76	65	77	70	74	112	111	113	133
52.5	58	67	104	113	132	121	128	174	178	209	225
57.5	102	100	127	181	195	237	238	220	253	336	391
62.5	134	185	213	222	258	345	411	427	411	428	531
67.5	140	195	257	271	328	442	572	672	621	618	635
Female											
	Period										
Age	1955.5	1960.5	1965.5	1970.5	1975.5	1980.5	1985.5	1990.5	1995.5	2000.5	2005.5
32.5	7	5	6	5	19	14	13	12	13	15	20
37.5	19	28	21	24	27	40	26	39	44	31	44
42.5	24	39	36	58	42	54	63	63	79	78	83
47.5	50	50	74	86	71	85	83	110	117	143	117
52.5	79	75	120	137	166	164	155	160	211	243	223
57.5	111	130	152	176	225	268	264	250	287	338	385
62.5	129	198	216	242	291	400	402	409	410	453	511
67.5	162	218	258	303	374	495	534	601	588	557	661

Table 2.3: *Person-years in 100 000 for colon cancer, by age and period.*

Male											
	Period										
Age	1955.5	1960.5	1965.5	1970.5	1975.5	1980.5	1985.5	1990.5	1995.5	2000.5	2005.5
32.5	6.65	5.83	5.15	5.11	6.38	8.05	7.87	8.11	8.36	8.91	8.63
37.5	6.45	6.56	5.76	5.12	5.11	6.37	8.05	7.91	8.10	8.41	9.01
42.5	6.17	6.37	6.48	5.71	5.09	5.08	6.34	8.01	7.85	8.09	8.44
47.5	5.59	6.07	6.26	6.38	5.62	5.02	5.01	6.25	7.90	7.77	8.05
52.5	5.08	5.45	5.91	6.08	6.21	5.47	4.89	4.88	6.11	7.75	7.66
57.5	4.53	4.88	5.21	5.63	5.81	5.92	5.23	4.68	4.71	5.93	7.52
62.5	3.66	4.23	4.54	4.82	5.22	5.39	5.50	4.86	4.39	4.47	5.65
67.5	2.82	3.31	3.76	4.01	4.27	4.63	4.78	4.90	4.39	4.02	4.12
Female											
	Period										
Age	1955.5	1960.5	1965.5	1970.5	1975.5	1980.5	1985.5	1990.5	1995.5	2000.5	2005.5
32.5	6.52	5.62	4.99	4.95	6.03	7.52	7.44	7.72	7.93	8.52	8.41
37.5	6.40	6.44	5.59	4.97	4.96	6.05	7.54	7.47	7.77	8.02	8.66
42.5	6.12	6.34	6.39	5.56	4.96	4.95	6.04	7.53	7.47	7.80	8.08
47.5	5.73	6.05	6.28	6.34	5.52	4.93	4.93	6.00	7.49	7.46	7.80
52.5	5.39	5.63	5.96	6.19	6.26	5.45	4.87	4.86	5.94	7.42	7.40
57.5	4.87	5.25	5.50	5.83	6.06	6.13	5.34	4.77	4.78	5.84	7.29
62.5	4.04	4.67	5.04	5.31	5.63	5.86	5.93	5.16	4.64	4.66	5.68
67.5	3.31	3.77	4.37	4.75	5.01	5.33	5.56	5.63	4.92	4.44	4.46

Table 2.4: *Incidence rates per 100 000 for colon cancer, by age and period.*

Male											
	Period										
Age	1955.5	1960.5	1965.5	1970.5	1975.5	1980.5	1985.5	1990.5	1995.5	2000.5	2005.5
32.5	1.20	2.23	2.33	1.17	1.88	1.86	3.18	1.85	1.44	1.68	2.67
37.5	2.64	3.96	2.95	3.51	2.94	4.55	2.73	3.79	2.10	3.45	3.22
42.5	5.83	5.18	6.02	5.61	5.70	7.48	8.05	7.74	8.79	7.17	8.77
47.5	7.87	8.73	12.15	10.19	13.70	13.95	14.76	17.93	14.06	14.54	16.53
52.5	11.41	12.29	17.60	18.57	21.27	22.13	26.19	35.63	29.15	26.98	29.39
57.5	22.52	20.50	24.38	32.13	33.59	40.02	45.54	47.05	53.72	56.71	51.97
62.5	36.64	43.72	46.96	46.01	49.41	64.02	74.70	87.85	93.55	95.78	93.94
67.5	49.57	58.91	68.29	67.63	76.81	95.42	119.55	137.03	141.61	153.74	154.04
Female											
	Period										
Age	1955.5	1960.5	1965.5	1970.5	1975.5	1980.5	1985.5	1990.5	1995.5	2000.5	2005.5
32.5	1.07	0.89	1.20	1.01	3.15	1.86	1.75	1.55	1.64	1.76	2.38
37.5	2.97	4.35	3.76	4.83	5.45	6.61	3.45	5.22	5.67	3.87	5.08
42.5	3.92	6.15	5.63	10.43	8.47	10.90	10.43	8.37	10.57	10.00	10.27
47.5	8.72	8.27	11.79	13.56	12.86	17.24	16.85	18.33	15.61	19.18	15.00
52.5	14.66	13.32	20.15	22.15	26.52	30.08	31.81	32.91	35.51	32.76	30.15
57.5	22.79	24.78	27.63	30.20	37.14	43.74	49.44	52.39	60.02	57.86	52.84
62.5	31.94	42.42	42.82	45.58	51.68	68.24	67.80	79.20	88.44	97.26	90.04
67.5	48.95	57.83	59.08	63.77	74.62	92.95	95.98	106.71	119.52	125.57	148.34

The tables of incidence rates are given in Table 2.4. Compared to the Lexis diagram given in section 2.2 age is given in ascending order in the tables. Thus the birth cohorts are given on the diagonal down and towards right, which is opposite to the Lexis diagram. If the incidence rates change simultaneously for all age groups for a specific birth cohort or period we say we have a cohort or period effect respectively. The intention of introducing explorative data analysis is to explore such features of the data. Thus it will be interesting to examine possible cohort or period effects for the cancer

sites in question. Fortunately we can easily obtain figures for examining both possible cohort and period effects in the so called CA- (rates vs. cohort by age) and PA- (rates vs. period by age) plots.

Figure 2.2 gives CA-plots (upper panel) and PA-plots (lower panel) for colon cancer by sex. The figures given on the left are for males and the figures on the right are for females. In the CA-plots the cohorts are given on the horizontal axis. Similarly in the PA-plots the periods (date of diagnosis) are given on the horizontal axis. For both plots the incidence rates per 100 000 are given on the vertical axis. For each age group the line represents the incidence rates over time. We expect the incidence rates to increase by age and time and this feature is captured in both the CA- and PA-plots for colon cancer. That is, we observe that the lines are higher the older the age group. Similarly the lines are higher for the latest compared to the earliest time periods for all age groups. For a specific birth cohort or period, we observe the incidence rates for all age groups simultaneously by following a vertical line in the CA- or PA-plot respectively. Thus by following a vertical line for the birth cohorts around WWII in the CA-plots we notice a decrease in the lines for almost all the age-groups for both sexes, which indicates that we have a birth cohort effect for those born around WWII. However it is not easy to observe a possible period effect for either males or females.

CA-plots are given for all cancer sites considered in this study in Appendix A. By concentrating the eye on the birth cohorts around WWII it might be possible to observe a transient reduction in the incidence rates for other sites as well. We should be careful however, not to over overinterpret the figures. The figures are discussed more closely in Chapter 5.

Even though the CA-plot imply that there might be a birth cohort effect for the cohorts born around WWII for colon cancer, statistical methods aid determining whether the trends are real or random. Since the dependent variable, the number of new cases, is a count, the model to be considered is the Poisson regression model, with age, period and cohort as covariates and log person-years as offset. This model is introduced in the following chapter.

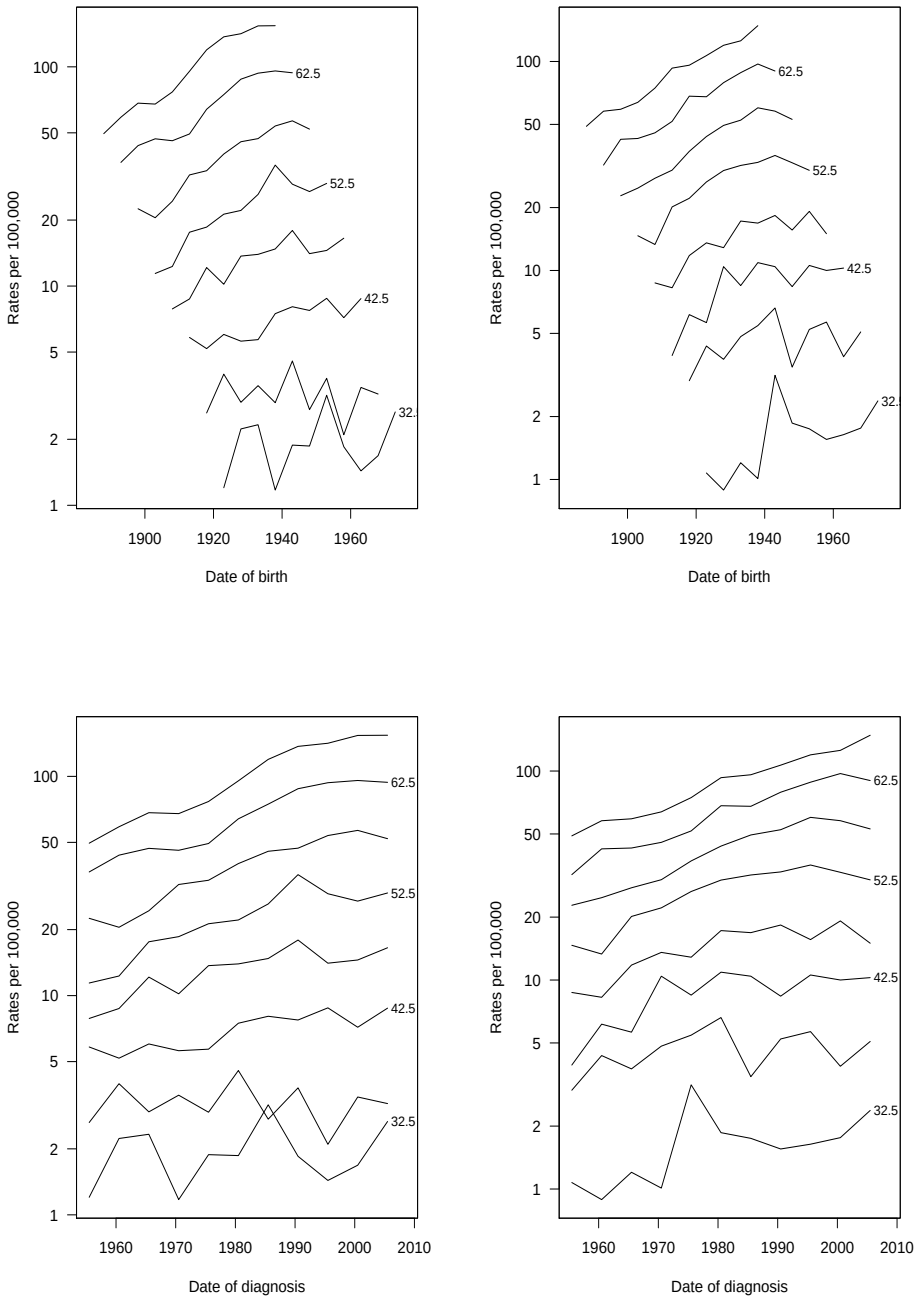


Figure 2.2: *CA- and PA-plots for colon cancer by sex. The CA-plots are in the upper panel and the PA-plots in the lower panel. The figures given on the left are for males and the figures on the right are for females.*

Chapter 3

Age-period-cohort model

The age-period-cohort model (apc model) is a well-known tool used by statisticians world wide when it comes to analysis of temporal patterns in disease data and will be introduced in this chapter. The apc model allows for measuring age, period and cohort effects simultaneously.

An estimator for the incidence rates for age group a and period p is defined as $\hat{r}_{ap} = \frac{d_{ap}}{Y_{ap}}$, where d_{ap} and Y_{ap} are given as the corresponding number of new cases and person-years. We consider the person-years to be non-random. The number of new cases, d_{ap}, are counts and we assume they are independent and Poisson distributed. Thus we assume $d_{ap} \sim Po(r_{ap}Y_{ap})$ where the rate r_{ap} is the expected number of cancer cases per person-year in age a and period p. We may consider a Poisson regression model where we implement the number of new cases, d_{ap}, as the response. For a Poisson regression model the mean $r_{ap}Y_{ap}$ of d_{ap} is explained in terms of the explanatory variables via an appropriate link, $g()$ (e.g. de Jong and Heller, 2008). To restrain the mean to be positive we consider the log-link. Then

$$g\left(E(d_{ap})\right) = \log E\left(d_{ap}\right) = \log(r_{ap}Y_{ap}) = \log r_{ap} + \log Y_{ap}.$$

In an age-period-cohort model we assume that $\log r_{ap}$ is a linear function of age, period and cohort effects, cf. below. The model may be fitted by the software **R** for Poisson regression by including $\log Y_{ap}$ as offset (see R Development Core Team, 2010).

From section 2.2 we have that cohort c is expressed by age group a and period p, that is $c = p - a$. Due to the linear dependency between the three covariates the model should be handled with care (Holford, 1991). In addition we should not trust that statistical models will provide definite answers and results for something as complex as trends in the number of new

cancer cases (Bray, 2005). Nevertheless when used with care and caution the apc model will aid to interpretation of the trends in incidence rates for the birth cohorts around WWII.

Before we introduce the full age-period-cohort model, we will introduce the so-called age, age-drift, and age-cohort and age-period models. The models can be seen as the hierarchy of models given in Figure 3.1 (Clayton and Schifflers, 1987a,b).

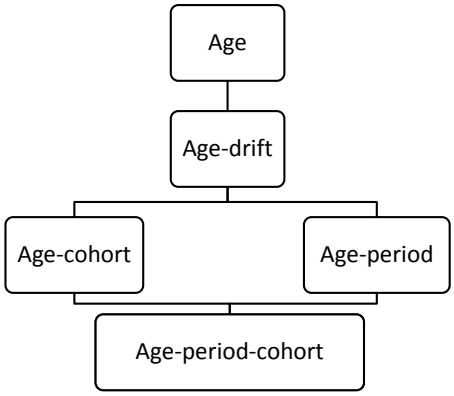


Figure 3.1: *Hierarchy of models introduced by Clayton and Schifflers.*

The term *drift* represents the average annual change in the rates over time (Bray, 2005) and will be discussed in section 3.2, where the age-drift model is introduced. The model considered further in this thesis is the full apc model. However, the apc model is the last model in the model-hierarchy and by introducing the other models first we will more easily understand the full apc model. We start by introducing the age model and work our way down the hierarchy of models.

Fortunately the function *apc.fit*, developed by Bendix Carstensen, in the Epi package in the software **R** (R Development Core Team, 2010) compute the age, period and cohort effects. Thus the function is used for all the models fitted throughout this chapter.

3.1 Age model

The age model is the simplest model included in the hierarchy of models given in Figure 3.1. As the name of the model implies the only covariate considered in this specific Poisson regression model is age. We use age as

a categorical covariate. With a log-link the rates can then be explained in terms of age by

$$\log(r_{ap}) = \log E(\hat{r}_{ap}) = \mu + \alpha_a \tag{3.1}$$

where μ is the rate for the reference group and where α_a measures the effect of age group a relative to the reference. Note that the estimated rates are presented visually as $e^{\hat{\mu}+\hat{\alpha}_a}$. As cancer rates always depend on age, the age model can be considered as the null hypothesis of no temporal variation (Clayton and Schifflers, 1987b, pg. 470).

3.2 Age-drift model

The second model suggested by Clayton and Schifflers is the age-drift model. Due to linear dependency between age, period and cohort, there is a linear variation over time which can be predicted by both the age-period and age-cohort model (Clayton and Schifflers, 1987a). This temporal variation can be considered as the drift, δ, and may be estimated by considering the following model

$$\log r_{ap} = \mu + \alpha_a + \delta \cdot j \tag{3.2}$$

where μ and α_a can be considered as in the age model. The drift is estimated by either specifying period or cohort as a continuous covariate, i.e. $j = p$ or $j = c$. The model will have the same estimated value for δ and the same fitted values of r_{ap} whether period or cohort is used to model the drift. However the age effects α_a will differ, and we cannot distinguish which of the two models represents the true age curve. That is, the reference will change depending on whichever of period or cohort is included in our model (Clayton and Schifflers, 1987a, pg. 462). As an example we consider the age effects estimated from the age-drift model for colon cancer, see Figure 3.2. The estimated effects are for considering cohort as a continuous variable. Similarly the dashed lines represents the estimated effects when considering period as a continuous variable. The figures to the left are the estimated effects for males and the figures to the right are the corresponding effects for females. From the figure we see that the estimated effects for age differ depending on whichever of period or cohort are given as the continuous covariate. Thus, the drift describes the temporal variation unattributable to specifically period or cohort influences.

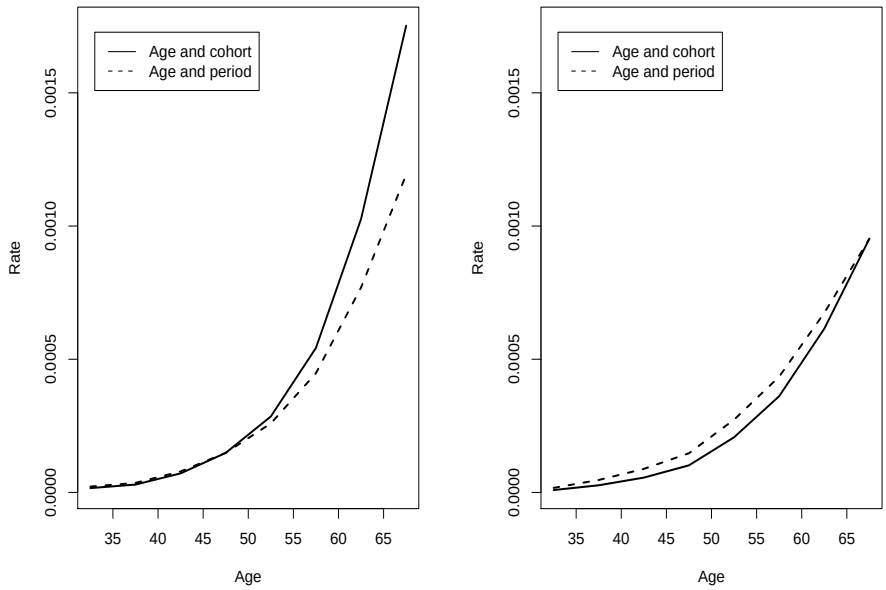


Figure 3.2: *Estimated age effects estimated from the age-drift model for colon cancer in Norway 1953-2007. Estimated effects for males are given on the left and on the right for females. The lines represent the estimated effects when including cohort as a continuous variable and the dashed lines on considering period as a continuous variable.*

The age-drift model is not of great interest by itself. However it is important to understand how the linear dependency between age, period and cohort influences the results. This will help us make valid interpretations of the result we obtain by using the full apc model later in this thesis.

3.3 Age-period and age-cohort models

In the hierarchy of models given in Figure 3.1, the next level is shared between the age-period and the age-cohort models. The models can be given as

$$\log r_{ap} = \mu + \alpha_a + \beta_p \tag{3.3}$$

or

$$\log r_{ap} = \mu + \alpha_a + \gamma_c \tag{3.4}$$

where μ and α_a are defined as above. Further β_p and γ_c are given as the period and cohort effect for period p and cohort c. The estimated rates will be presented visually as $e^{\hat{\mu}+\hat{\alpha}_a}$, similarly as for the age model. The estimated period and cohort effects will be presented visually as the relative risks, that is $e^{\hat{\beta}_p}$ and $e^{\hat{\gamma}_c}$. For the age-period model we assume no cohort effect, i.e. that the drift is allocated to period. Similarly for the age-cohort model, we assume no period effect (Clayton and Schifflers, 1987a). Choosing a reference cohort with relatively high number of new cases will make the fitted cell rates for the age-cohort model more reliable (Clayton and Schifflers, 1987a, pg. 460). Fortunately we have already excluded the youngest and oldest age groups and rely on the reference cohort to be chosen as a cohort with sufficient number of new cases.

As an example we consider the estimated effects from the age-period and age-cohort model for colon cancer. The estimated effects for males are given in Figure 3.3 and in Figure 3.4 for females. The estimated age effects for females look relatively similar. Although by closer examination we can see that they slightly differ. When it comes to the estimated cohort effects from the age-cohort model, we assume no non-linear period effect, and we see that all drift is allocated to cohort. From the age-cohort model a possible wartime effect is apparent for both sexes in the estimated effects for cohort. The estimated effects from the age-cohort and age-period model are only given as examples in this section. Due to the fact that these specific model are not of key interest, we will not discuss the results any further.

As there are less parameters used in the age-period model compared to the age-cohort model, which can be seen by the Lexis diagram given in section 2.2, it is not unlikely that the age-cohort model will have a better fit than the age-period model (Clayton and Schifflers, 1987a, pg. 466). However as the two models are not nested it is not straightforward to tell which, if any, model is better than the other (Clayton and Schifflers, 1987b, pg. 470). Thus we fit the full apc model and compare the age-cohort and the age-period model to the 3-factor-model. The apc model is the last model in the hierarchy of models suggested by Clayton and Schifflers and is introduced in the next section.

3.4 Age-period-cohort model

The full apc model measures the effect of age, period and cohort simultaneously. We are mainly interested in examining the trends in incidence rates for the birth cohorts in this thesis. However, by including period in the

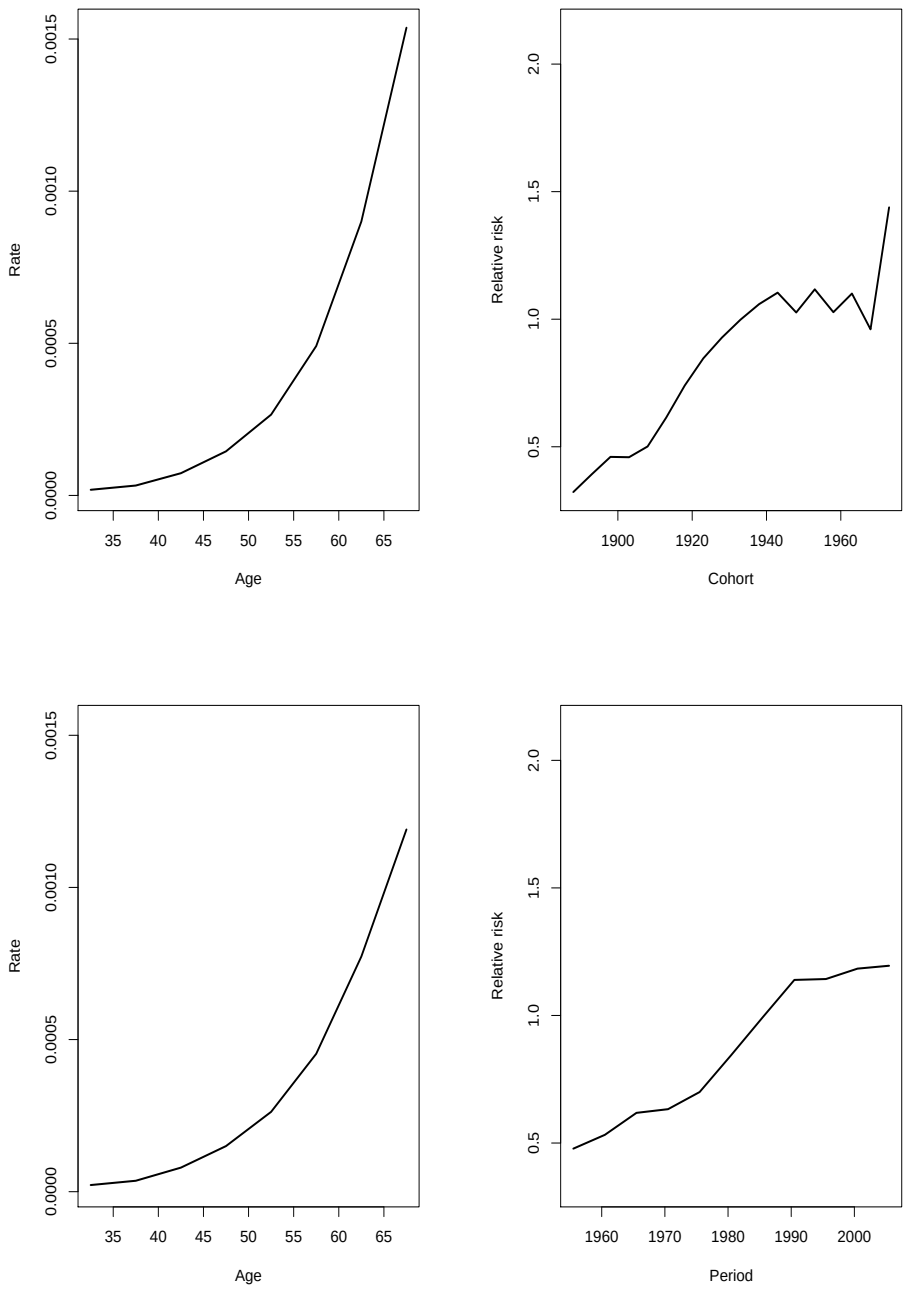


Figure 3.3: *Estimated effects for the age-cohort model are given in the upper panel and estimated effects for the age-period model are given in the lower panel. Estimated age effects are given on the left and estimated cohort and period effects are given on the right. The estimated effects are for colon cancer for males in Norway 1953-2007.*

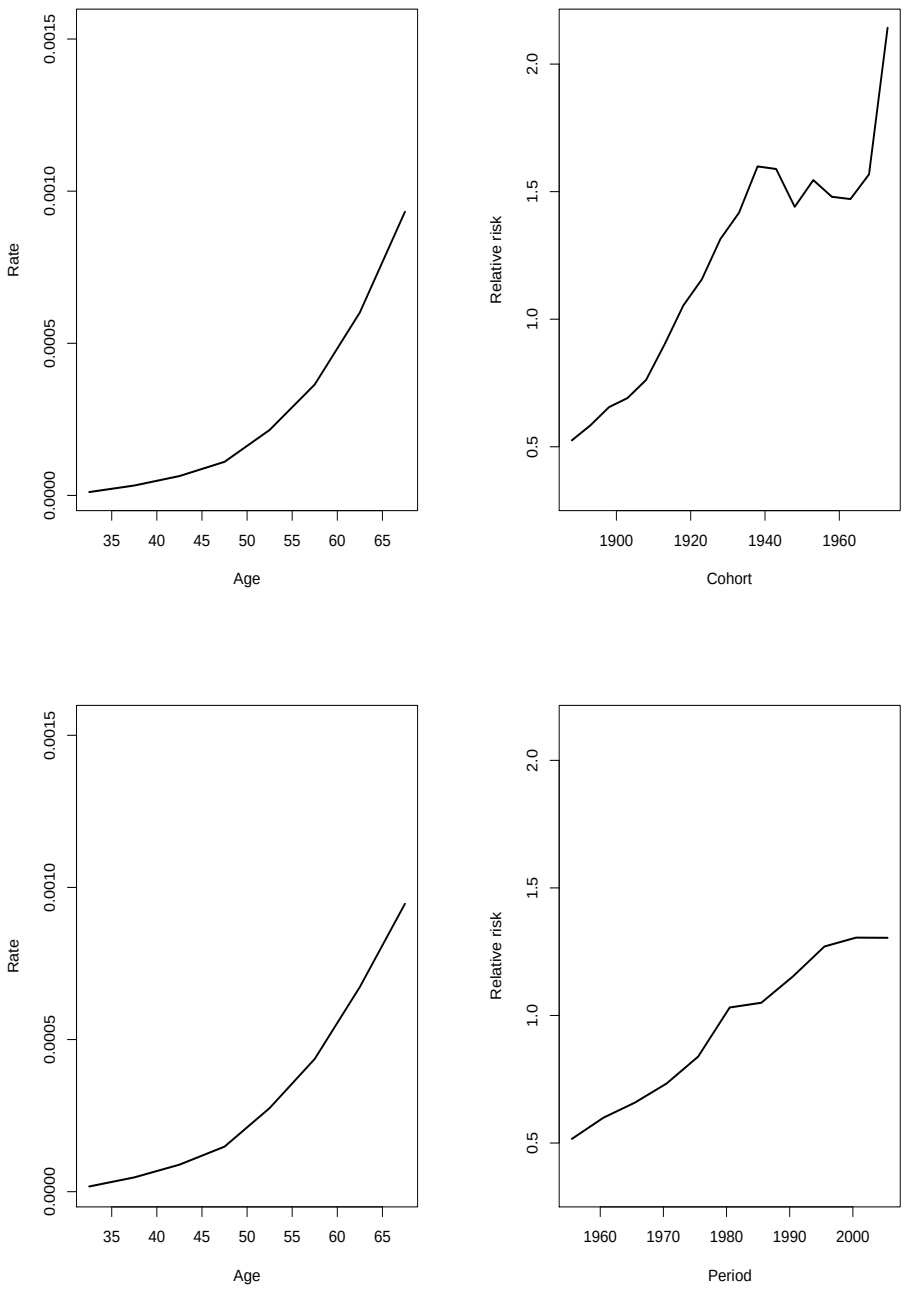


Figure 3.4: *Estimated effects for the age-cohort model are given in the upper panel and estimated effects for the age-period model are given in the lower panel. Estimated age effects are given on the left and estimated cohort and period effects are given on the right. The estimated effects are for colon cancer for females in Norway 1953-2007.*

model we are adjusting for non-linear period effects as well, which will make our interpretations more reliable. The Poisson regression model for the apc model can be given as

$$\log r_{ap} = \mu + \alpha_a + \beta_p + \gamma_c \tag{3.5}$$

see Clayton and Schifflers (1987b, pg. 472), where μ is given as the reference. Further we consider α_a, β_p and γ_c as the effect for age group a, period p and cohort c respectively. We would like to find out how well the models, given in the hierarchy in Figure 3.1, fit the data. Thus we include the analysis of deviance tables by sex. Basically an analysis of deviance table compares the models of interest to a saturated model, where a saturated model is a model with as many parameters as there are observations. Thus the saturated model fits perfectly (e.g. de Jong and Heller, 2008). We define the maximum possible log-likelihood for the saturated model as $\tilde{l}$ and as $\hat{l}$ for the model at interest. Further we define the *deviance* as

$$\Delta = 2(\tilde{l} - \hat{l})$$

that is the distance between the saturated and fitted model. A model that gives a good fit will have a log-likelihood value close to the log-likelihood value for the saturated model. Thus the smaller the deviance value, the better the fit. Further we have that

$$\Delta \sim \chi^2_{n-p},$$

i.e. the deviance is chi distributed with $n-p$ degrees of freedom. This yields if our models are adequate. For further details see for example de Jong and Heller (2008).

The analysis of deviance tables for colon cancer are given in Table 3.1 for males and in Table 3.2 for females.

Table 3.1: *Analysis of deviance for males experiencing colon cancer. Results are given for all models, including the age-drift model with both cohort and period given as a continuous variable.*

Model	Df	Δ	Change df	Change Δ	P($>\lvert\chi^2\rvert$)
Age	80	1445.35			
Age-drift	79	238.55	1	1206.81	0.000
Age-cohort	63	70.90	16	167.65	0.000
Age-period-cohort	54	53.74	9	17.16	0.050
Age-period	70	155.48	-16	-101.74	0.000
Age-drift	79	238.55	-9	-83.07	0.000

Table 3.2: *Analysis of deviance for females experiencing colon cancer. Results are given for all models, including the age-drift model with both cohort and period given as a continuous variable.*

Model	Df	Δ	Change df	Change Δ	P(> $\|\chi^2\|$)
Age	80	1401.06			
Age-drift	79	222.32	1	1178.74	0.000
Age-cohort	63	66.89	16	155.43	0.000
Age-period-cohort	54	46.95	9	19.94	0.020
Age-period	70	152.63	-16	-105.68	0.000
Age-drift	79	222.32	-9	-69.69	0.000

The first column in tables 3.1 and 3.2 represents the models given in Figure 3.1. The second and third column represents the degrees of freedom, $n - p$, and the deviance corresponding to the model given in the first column. The fourth and fifth column gives the change in degrees of freedom and deviance, except for the age model. If the models are not nested this does not make any statistical sense. As mentioned in section 3.3, the age-cohort and age-period model are not nested. Therefore they are both compared to the full apc model. The last column contain p-values for comparing the reduction in deviance for the row to the residuals. Thus we should consider the model(s) with a p-value higher than 5% or 1% significance level. From section 3.2, where the age-drift model is discussed, we know that the model will have the same fitted values for whichever of period or cohort are chosen to be included the model. Hence, as we can see from the analysis of deviance tables by sex, the age-drift model has the same deviance. Due to the difference in the number of parameters included in the two different models, which is discussed in section 3.3, we see from the tables that the age-cohort model has a better fit than the age-period model for both sexes. However, for both males and females, we see that the only model which gives a good fit is the apc model.

3.4.1 Holford's drift

As discussed above, we should be careful when interpreting the results from the apc model due to the linear dependency between the three factors age, period and cohort. Thus, we should find a way to extract the drift to make the interpretations easier. The usual constraints given for this model are $\alpha_a = 0$, $\beta_p = 0$ and $\gamma_c = 0$ for the first age group, period interval and cohort. However due to the linear dependency between age, period and cohort, these constraints are not sufficient. Thus an additional constraint is necessary (Heuer, 1997). However, as there exists no *a priori* information before the additional constraint is defined, this may lead to many different choices of the constraint. It can be given as $\gamma_C = 0$ or $\beta_{P-1} = \beta_P$ or

anything else. Thus the parameter estimates for age, period and cohort will depend on the specific restrictions used. However the models all obtain the same fitted values, regardless of the parameter estimates, and this is referred to as the problem of *non-identifiability*.

Holford (1991) figured that if we find the common features of all possible sets of allowed parameters, it will be possible to interpret the trends for age, period and cohort effect in a specific problem at hand. He suggests that we remove the overall linear trend (slope) and consider the remaining residuals, which can be interpreted as the curvature. Denote by $A,\ P$ and C the total number of age groups, periods and cohorts and introduce

$$\begin{aligned}
\alpha_a &= \left(a - \frac{A+1}{2}\right)\alpha_L + \phi_a \\
\beta_p &= \left(p - \frac{P+1}{2}\right)\beta_L + \phi_p \\
\gamma_c &= \left(c - \frac{C+1}{2}\right)\gamma_L + \phi_c
\end{aligned}$$

where $\alpha_L,\ \beta_L$ and γ_L represents the slope for age, period and cohort and where $\phi_a,\ \phi_p$ and ϕ_c represents the corresponding curvature (Bray, 2005, page 92). The relationship between the three covariates, $c = p - a$, leads to linear terms which are not identifiable. On the other hand the curvatures are identifiable.

Although the slopes may vary considerably for the various sets of parameters, due to the linear dependency, there are still limitations on the variations. Consider the linear terms for the three covariates. Then for any pair of numbers $(x,\ y)$ the linear combination

$$x\alpha_L + y\beta_L + (y - x)\gamma_L$$

is identifiable. As an example we consider $x = y = 1$, which shows that $\alpha_L + \beta_L$ is identifiable. Choosing $x = 0,\ y = 1$, we see that we may estimate the sum of the period and cohort $\beta_L + \gamma_L$, which will be denoted *Holford's drift* (e.g. Bray, 2005, page 92). Holford's drift is usually a good approximation to Clayton and Schifflers's interpretation of the drift, δ, given in (3.2). If we fix one of the slopes, $\alpha_L,\ \beta_L$ or γ_L, to a particular value, the two other slopes are determined. Thus the linear slopes are dependent of each other. Now consider an unknown constant v and define the three slopes

for an arbitrary model as

$$\begin{aligned} \alpha_L^* &= \alpha_L + v \\ \beta_L^* &= \beta_L - v \\ \gamma_L^* &= \gamma_L + v \end{aligned}$$

where α_L, β_L and γ_L are the true slopes. Our main interest in this thesis is examining the trends in the incidence rates for birth cohorts around WWII. Visual inspections of the estimated rates with all the drift placed in cohort will make it easier for us to spot a possible decrease in the incidence rates. More formally we will assume no period slope, i.e. $\beta_L = 0$. Thus $v = -\beta_L^*$ which gives $\alpha_L = \alpha_L^* + \beta_L^*$ and $\gamma_L = \beta_L^* + \gamma_L^*$, where we recognize $\beta_L^* + \gamma_L^*$ as Holford's drift. We will use Holford's interpretation of extracting the drift when estimating age, period and cohort estimates in this thesis. Further details of how to manage and interpret the apc model can be found in several different written documents such as Holford (1991) and Bray (2005).

As an example we consider the estimated age, period and cohort effects for colon cancer where we extract the drift by Holford's method. We will present two scenarios graphically, where we in the first scenario assume no period slope and place all the drift in cohort. In the second scenario we place all the drift in period, see Figure 3.5. The figures to the left are the estimated age, period and cohort effects for males and the figures to the right are the corresponding effects for females. As we can see estimated age effects by sex are slightly affected by the choice of where we put the drift. In addition, the estimated period and cohort effects are obviously affected depending on where drift is allocated.

As assumed a decrease in the birth cohorts around WWII for both males and females are present in Figure 3.5. We observe that it is easier to spot the decrease in the incidence rates for the birth cohorts for the estimates given by allocating all drift to cohort. As we are not particularly interested in the period effects they are not discussed in details here. However as for illustration, we see that the period effects differ depending on the choice in whichever of period or cohort we place the drift as is expected by the discussion given above.

3.5 Natural regression splines

When using the apc model it is most common to use data grouped by 5-year age and period intervals. An advantage of using wider time intervals is that the estimated effects are fairly smooth in graphic presentations, as for

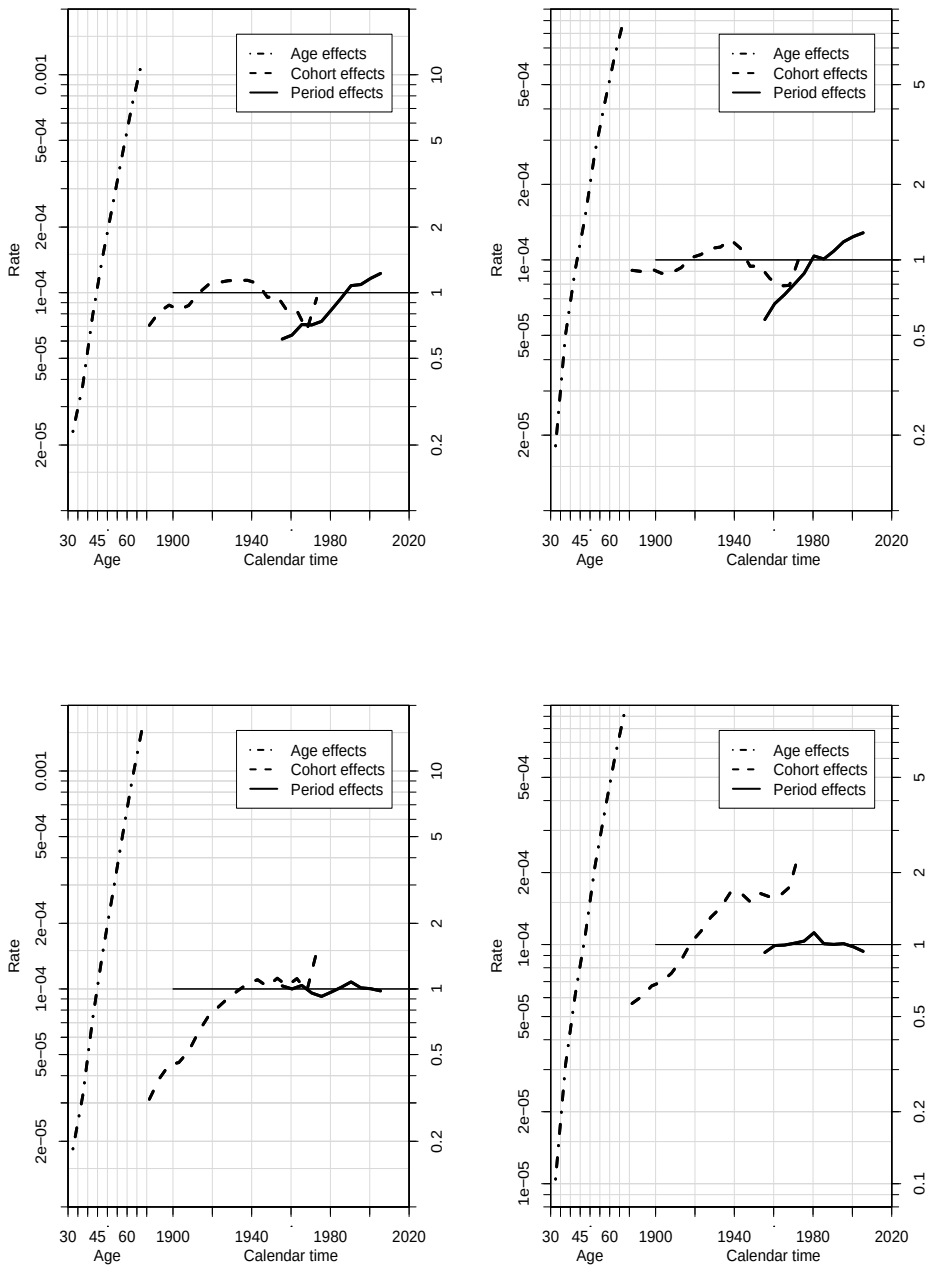


Figure 3.5: *Estimated age, period and cohort effects for colon cancer in Norway 1953-2007. The figures in the upper panel represents the estimated effect with all drift allocated to period. Similarly the figures in the lower panel represents the estimated effects with all drift allocated in cohort. The figures given on the left are for males and the figures on the right are for females.*

example in Figure 3.5. A disadvantage of grouping data in larger intervals is the loss of information, which can in many cases lead to an incomplete interpretation of data. The occupation period lasted for a total of five years and we hope to gain more information by using yearly data. Thus we introduce *splines*.

By using yearly data we may obtain estimated age, period and cohort effects from the apc model introduced above. However the curves of the estimated effects will not be as smooth as for the estimated effects when using 5-year age and period intervals. For our interpretations to be as reliable and accurate as possible, we would like to examine the estimated effects graphically. Splines are useful when considering yearly data in the apc model. Heuer (1997) gives a detailed description of how to include regression splines in the apc model where parts of the details will be presented in this section. Spline functions are well known in mathematical and thus statistical context and there exists a numerous number of spline functions. The area of application for spline functions are interpolation and smoothing, in which the latter is of particular interest.

For general regression splines we assume a time interval (a, b) partitioned in $m + 1$ subintervals. The subintervals are defined by m inner knots $\xi_1 < \xi_2 < \ldots < \xi_m$ and we define the outer knots as $a = \xi_0$ and $b = \xi_{m+1}$. In each subinterval we fit a polynomial of degree q (Heuer, 1997). To ensure that the polynomials for different subintervals are smoothly joined at the knots, and hence gives a smooth looking function, we assume that the piecewise polynomial functions are $q - 1$ times differentiable at the knots. As we will use splines in the context of the apc model, we have to consider the issue of non-identifiability, discussed in section 3.4. In addition the spline curves need to be stable in the tails. This is due to the low number of new cases for the earliest and latest age groups, periods and especially birth cohort groups. However, as we already omitted the youngest and oldest age groups this might not be crucial in our case. But we will still consider the spline function suggested by Heuer (1997). Thus we will consider natural regression splines with degree $q = 3$. More specifically these are defined as restricted cubic regression splines which are constrained to be linear in the tails, i.e. linearity is forced on the first interval (a, ξ_1) and on the last interval (ξ_m, b).

Heuer (1997) gives a thorough description of how we can integrate natural regression splines and the apc model. He introduces *B-splines* and shows how this particular spline function basis can be considered as natural regression splines by defining its degree $q = 3$ and restricting the function to be linear in the tails. Further he explains how we can manage the problem of non-identifiability and discuss how we can implement the method

introduced by Holford (1991), i.e. to separate constant, linear and nonlinear components, to the spline functions. Introducing spline functions and explaining how they can be applied in the apc model is a challenging task. The functions and their definitions are complicated. However Heuer (1997) has managed to present and describe the spline functions and their usage together with the apc model thoroughly and we refer to his work for substantial details regarding spline functions.

We give room for a short discussion concerning the choice of knots. Let $\tilde{N}$ be the number of observation years for either age, period or cohort. Then the number of inner knots is recommended as $m = [\tilde{N}/5]$, which returns the largest integer which does not exceeds its argument (see Heuer, 1997, pg. 169). We consider m to be the maximum number of inner knots when considering the apc model. As mentioned above the spline function is constrained to be linear at the boundaries. Thus the first and last inner knots (ξ_1 and ξ_m) need to have exceptional positions, see Heuer (1997, pg. 169) for details about the positions of the knots ξ_1 and ξ_m for age and period. However for cohort the first and last inner knots are defined to be $\xi_1 = 6$ and $\xi_m = C - 7$. Further the remaining $m - 2$ knots can be equally spread out in the interval (ξ_1, ξ_m), i.e. $\xi_i = \xi_1 + (i-1)\frac{\xi_m - \xi_1}{m-1}$ for $i = 1, \ldots, m$.

As we will see when using natural cubic splines for our data, the fluctuations may vary from cancer site to cancer site, and it is important not to overinterpret the plots. In the same way it is important that we choose the number of knots relatively large so we do not miss out any important trend changes, (Heuer, 1997, pg. 170). Even though Heuer (1997) has suggested the number of knots to use, he also recommends to vary the number of knots to find the number that fits the data in question best. It should be mentioned that he recommends at least four inner knots and not more than $m = [\tilde{N}/5]$ inner knots, which gives about one knot for every five years.

When using splines in practice, choosing the number of knots can be crucial to our interpretation of the results. In our case, since we are to compare different cancer sites, or at least if there is a specific year that has the main effect of the transient reduction of incidence in the birth cohorts, we should and will use the same number and position of the knots for all sites.

The age interval considered in this thesis is (30, 70), thus the number of knots should be $m = [\tilde{N}/5] = 8$. We hope to smooth the estimated curves as much as possible, without any loss of information, i.e. we have to be careful not to smooth out any possible non-linear effects in the curves. Fortunately the *apc.fit* function in the Epi package in the software R (R Development Core Team, 2010) allows us to easily integrate natural regression splines

in the apc model. In addition the function allows us to easily specify the number of knots we wish to include in the spline functions.

As an example we consider 4, 6, 8 and 10 knots for colon cancer by sex. In Figure 3.6 we observe that the variations in the estimated effects increase with the number of knots. For 4 knots the wartime effect seem to be smoothed away. It is not easy to distinguish between the figure using 6 and 8 knots. On the other hand, we expect the decrease to be around the same time for males and females and we can see that this will not be the case when we increase to 10 knots. Therefore we will hold on to Heuer's suggestion of using 8 knots for the best possible interpretation in this context. Compared to the estimated cohort effects for colon cancer when using 5-year data, see Figure 3.5, where all the drift is allocated to cohort we see that the overall trends in the rates are the same, which is as expected. Estimated cohort effects using natural splines with 8 knots for the sites in Table 2.1 are given in Appendix B and will be discussed in Chapter 5.

Comments

In this chapter we have introduced the apc model which is a well-known statistical model when it comes to analyzing disease data over time. Thus it is a suitable model when analyzing birth cohort rates for cancer data. We have discussed how we can use data given in 5-year age and period intervals by using age, period and cohort as factors in a Poisson regression model. We have also discussed how we can integrate spline functions and the apc model. As an attempt to recognize the expected transient reduction in the estimated cohort effects for colon cancer by sex, visual inspections are given. However, we would also like to have more formal conclusions of whether or not the wartime effect is significant. Thus, we introduce two statistical tests in the following chapter.

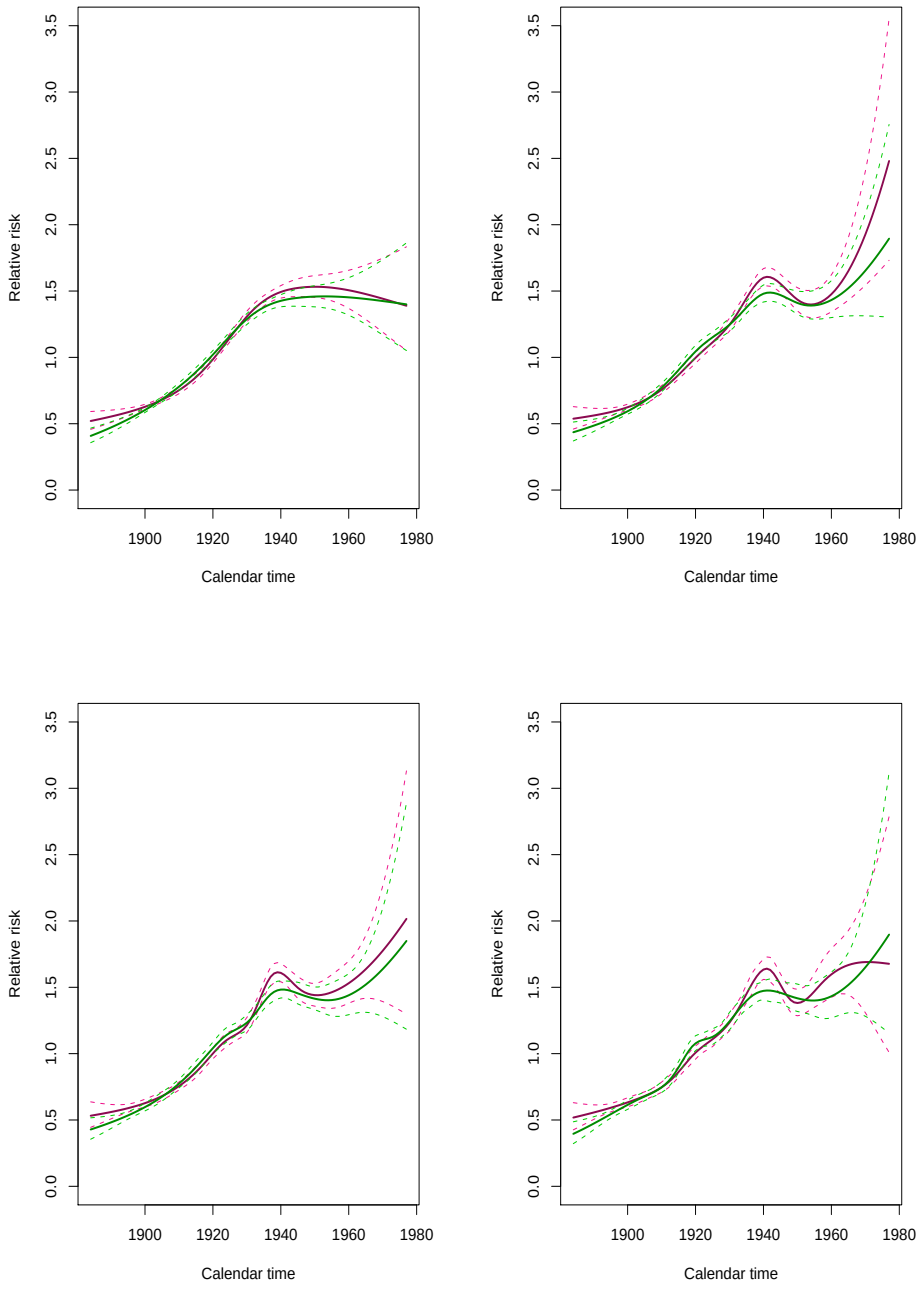


Figure 3.6: *Estimated cohort effects for colon cancer in Norway 1953-2007, using natural splines for yearly data. Green represents male and purple represents females. Notice the different scales on the y-axis for the different figures. The number of knots used is 4, 6, 8 and 10, respectively.*

Chapter 4

Testing cohort effects

In this chapter we will introduce two statistical tests, both introduced by the authors Tarone and Chu (1996, 2000). The first test can be considered as a generalization of the *second differences*. Thus a short recapitulation of second differences will be given before the test itself is introduced. Further the test is based on the estimated rates from the apc model where we compare the slopes between two time intervals and hope to identify any possible wartime effects. The first tests allows us to examine curvature for a coherent time interval as well. The second test is a non-parametric test which considers observed rates. However, Tarone and Chu (2000) suggest that the test is used adjunct to the results from the apc model. The second test is a generalization of the sign test. For both tests, examples will be given for colon, breast and testicular cancer.

4.1 Second differences

As mentioned several times above, the examination of period and birth cohort effects are not easy due to the linear dependency between the three factors age, period and cohort. However we still hope to identify any possible birth cohort effects by including both visual inspections and statistical test in our analysis. One idea, which is discussed in both Holford (1991) and Clayton and Schifflers (1987b), is to examine non-linear changes more closely. As will be discussed in section 4.2.1, the second differences are identifiable. Thus by examining the second differences we can see how a particular period or cohort deviates from the overall trend (e.g. Holford, 1991, pg. 22). We will only consider second differences for cohort effects in this thesis, however the principal ideas are the same for period effects. For

a specific cohort c consider the contrast

$$K_c = \gamma_{c+1} - 2\gamma_c + \gamma_{c-1} \tag{4.1}$$

which is defined by the second order differences of the cohort effects from the apc model. Thus which compares the difference between the change in the effects for birth cohort $c+1$ and c and birth cohort c and $c-1$. For further details I will refer to Holford (1991) and Clayton and Schifflers (1987b). However an example for colon cancer by sex is given in Figure 4.1, where we have used 5-year age and period intervals.

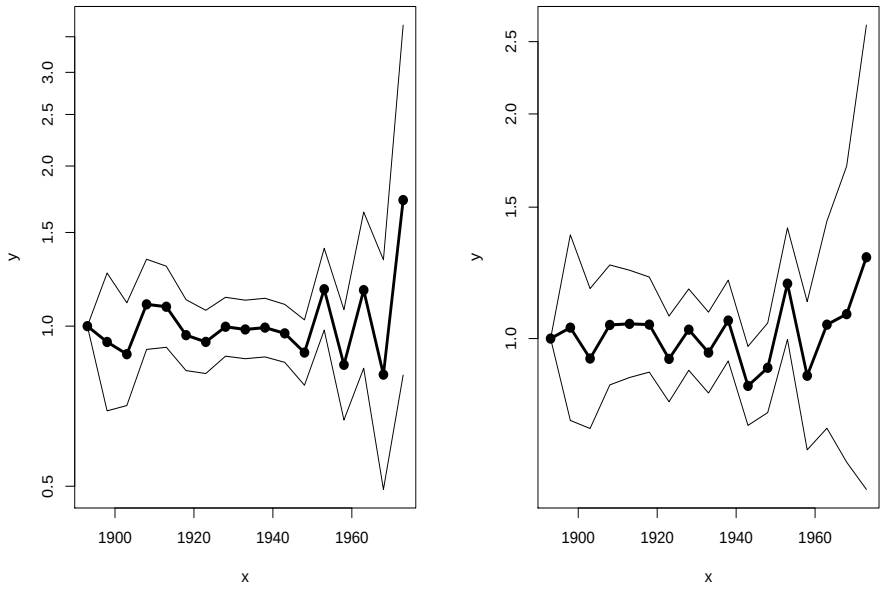


Figure 4.1: *Estimated second differences for birth cohorts using 5-year age and period intervals. The estimated effects are given for colon cancer by sex. The figure given on the left is for males and the figure on the right is for females.*

The plots given in Figure 4.1 are easily obtained by the function *contr.sec* implemented in the software **R**, see Appendix D. The figure given on the left is for males and the figure on the right is for females. We may say we have a decrease or increase in the incidence rates between adjacent birth cohorts if we observe a decrease or increase in the second differences given in the figure.

As we can see from Figure 4.1, due to random variation, it is not an easy task to observe any wartime effects for the birth cohorts. Colon cancer has

relatively high number of new cases compared to many of the other sites considered in this thesis. Thus it will be even more difficult to examine the figures for the other sites due to higher amount of random variation. It would be ideal to examine second differences for 1- and 2-year age and period intervals as well. However, these figures are subject to even more random variation compared Figure 4.1 where we have used 5-year age and period intervals. Therefore we will not examine second differences any further and we introduce the first formal test in the following section.

4.2 Testing for non-linear cohort effects

Extending the number of cohort effects included in the calculations in the previous section would make it possible to examine non-linear trends for wider intervals at the time, which may give more meaningful results and interpretations. The idea is adopted from a paper by Tarone and Chu (1996) and motivated with the fact that it is possible to identify the difference in slopes between two time intervals, without taking into consideration *how* we extract the drift from the apc model (see Tarone and Chu, 1996, Appendix 1). The number of cohort effects in each time interval is not set and depends on the problem at hand. Norway was occupied in a five year period during WWII. We therefore compare *cohort periods*, i.e. a time interval which consists of successive birth cohorts, with length of around five years. By using data with five year intervals for age and period we would only compare single cohort effects, thus using shorter time intervals will help us get a more detailed picture of the birth cohort effects before, during and after the occupation period, which is recommended by Tarone and Chu (1996) as well. Thus data with 1- and 2-year age and period intervals will be considered in this chapter. A description of the method is given in section 4.2.1.

By using the method of comparing slopes between two disjoint cohort periods, we assume that the slopes are linear. Now if the cohort effects are more like a second degree polynomial rather than linear, there is a way to capture the curvature when trying to identify a possible wartime effect. This method is described in section 4.2.2. Comparisons of the results based on change in linear trend and curvature for colon, breast and testicular cancer, are given in section 4.2.3.

4.2.1 Change in linear trend

Consider two cohort periods, C_1 and C_2, where η_1 and η_2 are the linear slopes of the two cohort periods. Thus we consider two cohort indices $c_A \leq c_B$,

such that the birth cohort effects are given by

$$\gamma_{c_1} = \theta_1 + \eta_1 c_1$$

for $c_1 \leq c_A$, and

$$\gamma_{c_2} = \theta_2 + \eta_2 c_2$$

for $c_2 \geq c_B$. In this case η_1 and η_2 are not identifiable, although the difference $\eta_2 - \eta_1$ is identifiable regardless of how the drift is extracted (Tarone and Chu, 1996).

The method introduced here will be considered for both data with 1- and 2-year age and period intervals, i.e. data with 2- and 4-year cohorts. The general method will be the same for both types of datasets, except some minor adjustment of the formulas depending on the data we use which we soon will see.

For data with 2-year age and period intervals Tarone and Chu (1996) consider the contrast

$$K = \gamma_{c_2+4} - \gamma_{c_2} - (\gamma_{c_1+4} - \gamma_{c_1}) \tag{4.2}$$

which compares the slope of two disjoint cohort periods with three consecutive births cohorts in each period. Another example of a contrast is

$$K = 3\gamma_{c_2+6} + \gamma_{c_2+4} - \gamma_{c_2+2} - 3\gamma_{c_2} - (3\gamma_{c_1+6} + \gamma_{c_1+4} - \gamma_{c_1+2} - 3\gamma_{c_1}) \tag{4.3}$$

with four consecutive births cohorts in each cohort period. For yearly data we consider similar contrasts where the general formula will be defined shortly.

More generally we consider contrasts of the form

$$K = s'\gamma, \tag{4.4}$$

where s and γ are given as vectors of weights and birth cohort effects. Let $\hat{\gamma}$ be the vector of estimated cohort effects from the apc model and introduce the estimated contrast $\hat{K} = s'\hat{\gamma}$. Then the variance of the estimated contrast $\hat{K}$, $\hat{\sigma}_K^2$, is estimated as $s'\hat{V}_\gamma s$, where $\hat{V}_\gamma$ denotes the estimated covariance matrix of $\hat{\gamma}$.

Furthermore we want to test the null hypotheses $H_0 : K = 0$ vs. the alternative $H_a : K \neq 0$. To this end we may use the test statistic

$$z = \frac{\hat{K}}{\hat{\sigma}_K}, \tag{4.5}$$

which is approximately standard normal under H_0. Note that the weights s may be scaled differently than the weights defined by Tarone and Chu (1996) without changing the value of the test statistic z. Using standardized weights as described below will restrain the estimated contrast $\hat{K}$ to have the same unit regardless of the number of cohort effects we include in each period and whether we use 1- or 2-year age and period intervals. Thus it will be easier to compare the value of the estimated contrasts in different scenarios. For simplicity we use the notation for the estimated contrast as $\hat{K} = s'\hat{\gamma}$ regardless of the scaling of weights we use.

The standardized weights can be justified by an argument using least squares. Thus a short recapitulation of the least squares method will help the reader understand how the birth cohort effects and weights are defined. Now assume $J+1$ data points (y_0, x_0), ..., (y_J, x_J), and fit a straight line $\theta + \eta x$ for the y_j's. The least square estimate for the slope is:

$$\hat{\eta} = \frac{\sum_{j=0}^{J}(x_j - \bar{x})(y_j - \bar{y})}{\sum_{j=0}^{J}(x_j - \bar{x})^2} = \frac{1}{M}\sum_{j=0}^{J}(x_j - \bar{x})y_j \tag{4.6}$$

where

$$M = \sum_{j=0}^{J}(x_j - \bar{x})^2.$$

Now for a specific cohort c, let $y_j = \hat{\gamma}_{c+j}$ and $y_j = \hat{\gamma}_{c+2j}$ for 1- and 2-year age and period intervals with $j = 0, 1, \ldots, J$. Further consider $x_j = j$ and $x_j = 2j$ for 1- and 2-year age and period intervals respectively. A more precise definition would be to let $x_j = c + j$ and $x_j = c + 2j$. However since the results will not be affected by this we will for simplicity define the x_j's as the former vectors.

As an example we consider the estimated birth cohort effects for a general period when considering 2-year data with four birth cohorts in each cohort period (cf. (4.3)) as

$$\hat{\gamma}_c,\ \hat{\gamma}_{c+2},\ \hat{\gamma}_{c+4},\ \hat{\gamma}_{c+6}. \tag{4.7}$$

Now assume we wish to find the estimated slope for the birth cohort estimates which is given by

$$\hat{\eta} = \frac{1}{M}\sum_{j=0}^{3}(2j-3)\ \hat{\gamma}_{c+2j}$$

where $M = \sum_{j=0}^{3}(2j-3)^2 = (-3)^2 + (-1)^2 + (1)^2 + (3)^2 = 20$. Thus we obtain $\hat{\eta} = \frac{1}{20}(-3\gamma_c - \gamma_{c+2} + \gamma_{c+4} + 3\gamma_{c+6})$. Except for the scaling $\frac{1}{20}$, this gives rise to the contrast (4.3).

Now for a more general statement consider $J+1$ estimated cohort rates included in each period. Then the estimated birth cohorts rates are given as

$$\hat{\gamma}_c,\ \hat{\gamma}_{c+1},\ \ldots,\ \hat{\gamma}_{c+J} \tag{4.8}$$

and

$$\hat{\gamma}_c,\ \hat{\gamma}_{c+2},\ \ldots,\ \hat{\gamma}_{c+2J} \tag{4.9}$$

for 1- and 2-year age and period intervals. The estimated slope for yearly data are then given as

$$\hat{\eta} = \frac{1}{M}\sum_{j=0}^{J}\left(j - \frac{J}{2}\right)\ \hat{\gamma}_{c+j}$$

where

$$M = \sum_{j=0}^{J}\left(j - \frac{J}{2}\right)^2$$

For 2-year age and period intervals the slope is given as

$$\hat{\eta} = \frac{1}{M}\sum_{j=0}^{J}(2j - J)\ \hat{\gamma}_{c+2j}$$

where

$$M = \sum_{j=0}^{J}(2j - J)^2$$

By using simple calculations and rules for series of sequences it is possible to find explicit formulas for M, although I will not do the actual calculations here.

So far we have considered two disjoint cohort periods. There might also be cases when the two cohort periods overlap, i.e. period C_2 starts where period C_1 ends. When this is the case s will slightly change. As an example consider two cohort effects in each period (cf. (4.2)) and let C be the cohort when $C_1 = C_2$ which leads to the contrast

$$K = \gamma_{c+4} - 2\gamma_c + \gamma_{c-4}.$$

which is exactly the second differences given in (4.1). Similarly for (4.3)

$$K = 3\gamma_{c+6} + \gamma_{c+4} - \gamma_{c+2} - 6\gamma_c - \gamma_{c-2} + \gamma_{c-4} + 3\gamma_{c-6}. \tag{4.10}$$

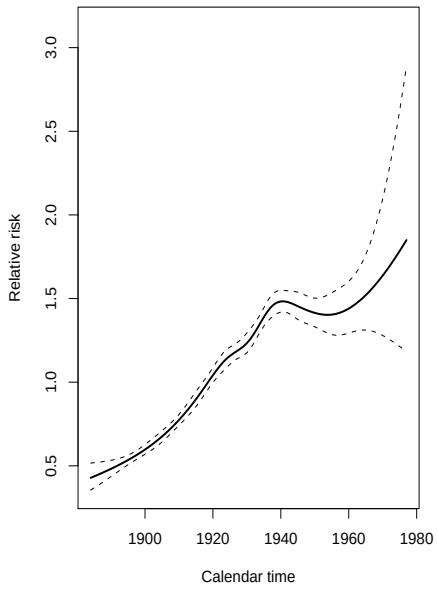
Note that M will have the same value as if the two cohort periods were disjoint.

For illustration examples are given for colon cancer by sex. Even though there are many different cohort periods we can examine, we will start by comparing the pre-war and occupation periods, see Table 4.1. The reader should keep in mind that the number of estimated cohort effects included in each cohort period will change depending on the dataset we use.

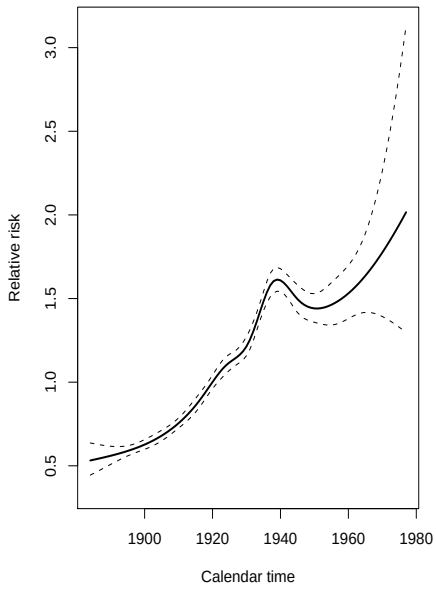
Table 4.1: *Estimated contrast $\hat{K}_i$ for colon cancer where $i = 1, 2$ represents 1- and 2-year age and period intervals respectively. Corresponding p-values are given to the right of the contrast. The two cohort periods under consideration are 1936-42 vs. 1942-48 for males and 1932-38 vs. 1938-44 for females. For yearly data for females we had to adjust the lower age group from 30 to 35 years, because of the small number of new cases for the youngest age groups. Standard deviations of the estimated contrasts are given in parentheses.*

	$\hat{K}_1$	p	$\hat{K}_2$	p
Male	-0.068 (0.021)	0.001	-0.055 (0.018)	0.002
Female	-0.063 (0.018)	0.000	-0.051 (0.015)	0.001

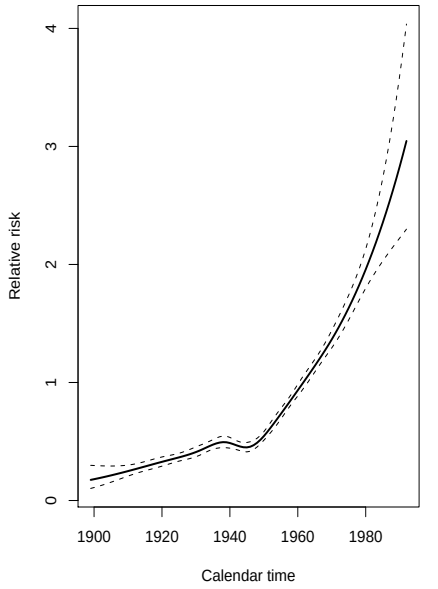
From Table 4.1 we see that the value of $\hat{K}$ slightly differ, depending on the dataset we use, even if the cohort periods under consideration are the same. We might think that the values should be similar since the motivation for the method introduced by Tarone and Chu (1996) is that the difference in slope, $\eta_2 - \eta_1$, is identifiable. However we should keep in mind that the data are divided in different time intervals depending on the dataset we use. The standard deviations (given in parentheses) are smaller for 2-year age and period intervals which is not unexpected as the random variation decrease by higher number of new cases in each cell. Estimated birth cohort effects for colon, breast and testicular cancer for 2-year data are given in Figure 4.2. The estimated effects are obtained using linear regression splines with eight knots where the figures given on the left are for males and the figures on the right are for females. We see that the reduction in the incidence rates are a little earlier for females than for males for those experiencing colon cancer. Thus the period, 1936-42 vs. 1942-48 for males and 1932-38 vs. 1938-44 for females, where chosen for this purpose. Apparently the linear slopes for both sexes are significantly decreasing, i.e. a negative contrast for both datasets, when comparing the cohort periods before and during the occupation period for Norway.



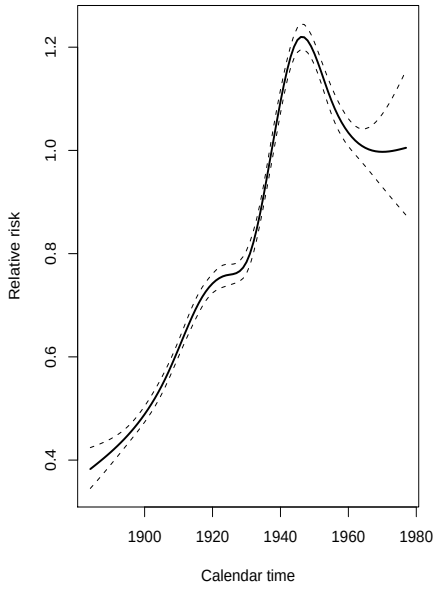
(a) Colon, male



(b) Colon, female



(c) Testis



(d) Breast

Figure 4.2: *Estimated birth cohorts for 2-year age and period intervals for colon, breast and testicular cancer in Norway 1953-2007. The estimated effects are obtained by using natural splines with eight knots. The drift is allocated in cohort and extracted by the method of Holford.*

4.2.2 Change in curvature

Figure 4.2 indicates that there may be more like a curvature than a change in slopes. A possible drawback of the method introduced in the previous section might be that it only considers change in linear trend. Modeling curvature can easily be obtained by the use of orthogonal polynomials. As an example consider the following estimated birth cohort effects for 2-year age and period intervals:

$$\hat{\gamma}_{c-6},\ \hat{\gamma}_{c-4},\ \hat{\gamma}_{c-2},\ \hat{\gamma}_{c},\ \hat{\gamma}_{c+2},\ \hat{\gamma}_{c+4},\ \hat{\gamma}_{c+6}$$

corresponding to (4.3) with one coherent time interval instead of two disjoint cohort periods. We may fit a second degree polynomial to the estimated rates given above, where the polynomial can be defined as

$$\gamma_{c+2j} = \theta v_{0j} + \eta_1 v_{1j} + \eta_2 v_{2j} \tag{4.11}$$

with $j = 0,\ \pm1,\ \pm2,\ \pm3$ where $v_{0j} = 1$, $v_{1j} = 2j$ and $v_{2j} = -4 + j^2$. Note that $\sum_{j=-3}^{3} v_{1j} = \sum_{j=-3}^{3} v_{2j} = 0$ and $\sum_{j=-3}^{3} v_{1j}v_{2j} = 0$. Thus the vectors $v_0,\ v_1$ and v_2 (with elements $v_{0j},\ v_{1j},\ v_{2j}$) are orthogonal. Now that we have defined an orthogonal set of vectors, the estimated η_2 from (4.11) will be the same as for $\gamma_{c+2j} = \theta v_{0j} + \eta_2 v_{2j}$, i.e. where we have removed the first-degree term. Thus an estimate of η_2 in (4.11) is obtained similarly as (4.6), that is

$$\begin{aligned}\hat{\eta}_2 &= \frac{1}{M}\sum_{j=-3}^{3}(v_{2j} - \bar{v}_2)\hat{\gamma}_{c+2j} \\ &= \frac{1}{M}(5\hat{\gamma}_{c+6} - 3\hat{\gamma}_{c+2} - 4\hat{\gamma}_{c} - 3\hat{\gamma}_{c-2} + 5\hat{\gamma}_{c-6})\end{aligned}$$

where $M = \sum_{-3}^{3}(v_{2j} - \bar{v}_2)^2 = 84$. It is common to standardize the weights and obtain an orthonormal set of vectors with length 1, i.e. $u_0 = v_0\frac{1}{\sqrt{7}}$, $u_1 = v_1\frac{1}{\sqrt{112}}$ and $u_2 = v_2\frac{1}{\sqrt{84}}$. The orthogonal vectors v_0, v_1 and v_2 are then replaced with the orthonormal vectors u_0, u_1 and u_2. However, for the formula given here, where we divide on M, the results obtained are the same as if we would use the orthonormal set of vectors. For yearly data spanned over the same cohort period, we consider the estimated effects:

$$\hat{\gamma}_{c-6},\ \hat{\gamma}_{c-5},\ \hat{\gamma}_{c-4},\ \hat{\gamma}_{c-3},\ \hat{\gamma}_{c-2},\ \hat{\gamma}_{c-1},\ \hat{\gamma}_{c},\ \hat{\gamma}_{c+1},\ \hat{\gamma}_{c+2},\ \hat{\gamma}_{c+3},\ \hat{\gamma}_{c+4},\ \hat{\gamma}_{c+5},\ \hat{\gamma}_{c+6}.$$

Similarly as for 2-year age and period intervals the polynomial can be defined as

$$\gamma_{c+j} = \theta v_{0j} + \eta_1 v_{1j} + \eta_2 v_{2j}$$

for $j = 0,\ \pm1,\ \pm2,\ \pm3\ \pm4,\ \pm5,\ \pm6$ and $v_{0j} = 1$, $v_{1j} = j$ and $v_{2j} = j^2 - 14$. Thus with $M = \sum_{j=-6}^{6}(v_{2j} - \bar{v}_2)^2 = 2002$ we obtain

$$\begin{aligned}\hat{\eta}_2 &= \frac{1}{2002}\sum_{j=-6}^{6}(v_{2j} - \bar{v}_2)\hat{\gamma}_{c+j} \\ &= \frac{1}{2002}(22\hat{\gamma}_{c+6} + 11\hat{\gamma}_{c+5} + 2\hat{\gamma}_{c+4} - 5\hat{\gamma}_{c+3} - 10\hat{\gamma}_{c+2} - 13\hat{\gamma}_{c+1} - 14\hat{\gamma}_c \\ &\quad -13\hat{\gamma}_{c-1} - 10\hat{\gamma}_{c-2} - 5\hat{\gamma}_{c-3} + 2\hat{\gamma}_{c-4} + 11\hat{\gamma}_{c-5} + 22\hat{\gamma}_{c-6})\end{aligned}$$

Further details of the method of orthogonal polynomials are somewhat tedious and given in Appendix C.

As mentioned above we consider one coherent cohort period when modeling curvature using orthogonal polynomials. However to make comparisons of change in linear trend and curvature as simple as possible we will refer to the cohort period as two separate cohort periods where period C_1 ends where period C_2 begins. Nevertheless it is necessary to define the orthogonal contrast as K_o so we do not confuse it with K defined in (4.4).

Similarly as Table 4.1 where we calculated K for colon cancer, we will here calculate K_o for 1- and 2-year age and period intervals, see Table 4.2.

Table 4.2: *Estimated contrast $\hat{K}_{o_i}$ for colon cancer where $i = 1,\ 2$ represents 1- and 2-year age and period intervals respectively. Corresponding p-values are given to the right of the contrast. The two cohort periods under consideration are 1936-42 vs. 1942-48 for males and 1932-38 vs. 1938-44 for females. For yearly data for females we had to adjust the lower age group from 30 to 35 years, because of the small number of new cases for the youngest age groups. Standard deviations of the estimated contrasts are given in parentheses.*

	$\hat{K}_{o_1}$	p	$\hat{K}_{o_2}$	p
Male	-0.224 (0.07)	0.001	-0.139 (0.049)	0.005
Female	-0.222 (0.06)	0.000	-0.151 (0.042)	0.000

By examining the curvature we conclude with the same as for change in linear trend, see section 4.2.1, i.e. the decrease of the cohort estimates during the occupation period compared to the time interval before the war is significant for both sexes, independent of the dataset we use. Similarly the standard deviations are smaller when considering 2-year age and period intervals compared the results obtained from yearly data.

4.2.3 Comparison of the change in linear and curvature

The cohort periods that are of interest to compare are the cohort periods before and during the occupation period, where we are looking for a significant decrease in the estimated birth cohort effects. Similarly we compare the cohort periods during and after the occupation period, where we may look for a significant increase. More formally, the estimated contrast is negative if a decrease is observed in the estimated effects. Similarly the estimated contrast is positive if an increase is observed. As mentioned in section 4.2.2 we consider one coherent cohort period when examining curvature. However we will for simplicity refer to it as two separate cohort periods. We will try both methods, i.e. change in linear trend and curvature, for 1- and 2-year age and period intervals for colon, breast and testicular cancer. Due to the relative large numbers of tests we perform we choose a significance level of 1%; more formally we could have considered a Bonferroni correction.

Table 4.3: *Estimated contrast $\hat{K}$ and $\hat{K}_o$ for colon cancer using yearly data. Corresponding p-values are given to the right of the contrast. The standard deviations of $\hat{K}$ and $\hat{K}_o$ are given in parentheses. Estimated contrasts significant on a 1% level are marked with $*$. Due to the small number of new cases for the youngest age groups for females, we had to adjust the lower age group from 30 to 35 years.*

Cohort period	$\hat{K}$	p	$\hat{K}_o$	p_o
		Male		
1932-38 vs. 1938-44	0.027 (0.018)	0.137	0.098 (0.060)	0.100
1934-40 vs. 1940-46	0.009 (0.020)	0.651	−0.085 (0.066)	0.196
1936-42 vs. 1942-48	−0.068* (0.021)	0.001	−0.224* (0.070)	0.001
1938-44 vs. 1944-50	−0.053 (0.021)	0.012	−0.082 (0.073)	0.262
1940-46 vs. 1946-52	0.082* (0.025)	0.001	0.174 (0.080)	0.030
1942-48 vs. 1948-54	0.094* (0.028)	0.001	0.245* (0.091)	0.007
1944-50 vs. 1950-56	−0.024 (0.031)	0.431	−0.074 (0.107)	0.490
		Female		
1932-38 vs. 1938-44	−0.063* (0.018)	0.000	−0.222* (0.060)	0.000
1934-40 vs. 1940-46	−0.043 (0.019)	0.025	−0.218* (0.063)	0.001
1936-42 vs. 1942-48	−0.023 (0.020)	0.252	−0.058 (0.066)	0.381
1938-44 vs. 1944-50	0.003 (0.022)	0.897	−0.024 (0.075)	0.746
1940-46 vs. 1946-52	0.053 (0.024)	0.023	0.173 (0.080)	0.030
1942-48 vs. 1948-54	0.071* (0.026)	0.006	0.298* (0.088)	0.001
1944-50 vs. 1950-56	0.010 (0.031)	0.741	−0.058 (0.104)	0.575

Table 4.4: *Estimated contrast $\hat{K}$ and $\hat{K}_o$ for colon cancer using 2-year age and period intervals. Corresponding p-values are given to the right of the contrast. The standard deviations of $\hat{K}$ and $\hat{K}_o$ are given in parentheses. Estimated contrasts significant on a 1% level are marked with $*$.*

Cohort period	$\hat{K}$		p	$\hat{K}_o$		p_o
			Male			
1932-38 vs. 1938-44	0.026	(0.016)	0.095	0.089	(0.040)	0.035
1934-40 vs. 1940-46	0.007	(0.018)	0.691	−0.063	(0.046)	0.176
1936-42 vs. 1942-48	−0.055*	(0.018)	0.002	−0.139*	(0.049)	0.005
1938-44 vs. 1944-50	−0.056*	(0.018)	0.002	−0.044	(0.052)	0.393
1940-46 vs. 1946-52	0.056*	(0.021)	0.008	0.088	(0.057)	0.124
1942-48 vs. 1948-54	0.070*	(0.024)	0.004	0.129	(0.065)	0.048
1944-50 vs. 1950-56	−0.001	(0.027)	0.957	0.034	(0.073)	0.644
			Female			
1932-38 vs. 1938-44	−0.051*	(0.015)	0.001	−0.151*	(0.042)	0.000
1934-40 vs. 1940-46	−0.046*	(0.017)	0.006	−0.158*	(0.044)	0.000
1936-42 vs. 1942-48	−0.039	(0.017)	0.023	−0.072	(0.047)	0.126
1938-44 vs. 1944-50	0.008	(0.019)	0.680	0.006	(0.052)	0.904
1940-46 vs. 1946-52	0.055*	(0.020)	0.007	0.142*	(0.055)	0.010
1942-48 vs. 1948-54	0.056	(0.023)	0.013	0.193*	(0.061)	0.002
1944-50 vs. 1950-56	0.006	(0.027)	0.831	−0.046	(0.073)	0.529

For colon cancer, see Table 4.3 and Table 4.4, the standard deviations, given in parentheses, are smaller for the contrasts $\hat{K}$ compared to the standard deviations for the contrasts $\hat{K}_o$ for both sexes. Similarly the standard deviations are smaller for data with 2-year age and period intervals compared to yearly data. When it comes to the p-values, there does not seem to be a specific pattern. However we should mention that there are more significant estimated contrasts in Table 4.4, i.e. where data with 2-year age and period intervals have been used, compared to Table 4.3. However as an overall summary the results seem to be similar for 1- and 2-year age and period intervals. There seem to be a significant wartime effect for both sexes, which is expected since the same conclusion have been reached in earlier studies, see Svensson et al. (2002).

Even though this is a progress in an attempt to find a possible wartime effect, similar tables should be obtained for breast and testicular cancer. This is due to the fact that earlier studies have concluded with significant wartime effects for these particular sites as well, see Tretli and Gaard (1996) and Wanderås et al. (1995). Thus we should look for reductions in the can-

cer rates for those born just before and during the war and in addition for the women experiencing puberty during the occupation period for breast cancer. We assume that the years of puberty are mainly for girls between ten and fifteen years old which can be considered as the birth cohorts around 1925-1935. The results are briefly discussed here. For yearly data we found a significant increase, that is we found a positive contrast, for both K and K_o for cohort period 1926-32 vs. 1932-38 on a 1% significance level. For cohort period 1924-1930 vs. 1930-36 and 1928-1934 vs. 1934-40 we found a significant increase in the curvature, that is the contrast K_o, for the same significance level. The results where similar for 2-year age and period intervals. Even though we where not able to capture a decrease in the birth cohort rates for those experiencing puberty during the occupation period, compared to those who experienced puberty before, we where able to capture a increase for those experiencing puberty during compared to after the occupation period. However as we can see from the figure for the estimated effects for breast cancer in Appendix B, the transient reduction for those experiencing puberty under WWII is relatively small, and this might be the reason why we are not able to capture any significant decrease. Thus we will not discuss this any further and we will now consider comparing the birth cohort born around WWII.

Tables 4.5 and 4.6 give the results for breast cancer for females and tables 4.7 and 4.8 for testicular cancer. As we can see from the tables the standard deviations of the estimated contrasts in the different sites and datasets are very similar. For instance, the values are around 0.009 for $\hat{K}$ for breast cancer for yearly data. However it should be noted that they are not similar by closer inspection.

Table 4.5: *Estimated contrast $\hat{K}$ and $\hat{K}_o$ for breast cancer for females using yearly data. Corresponding p-values are given to the right of the contrast. The standard deviations of $\hat{K}$ and $\hat{K}_o$ are given in parentheses. Estimated contrasts significant on a 1% level are marked with $*$.*

Cohort period	$\hat{K}$	p	$\hat{K}_o$	p_o
1932-38 vs. 1938-44	-0.019 (0.009)	0.039	-0.089^* (0.030)	0.003
1934-40 vs. 1940-46	-0.013 (0.009)	0.142	-0.062 (0.029)	0.030
1936-42 vs. 1942-48	-0.050^* (0.009)	0.000	-0.218^* (0.030)	0.000
1938-44 vs. 1944-50	-0.049^* (0.009)	0.000	-0.181^* (0.029)	0.000
1940-46 vs. 1946-52	-0.024^* (0.009)	0.006	-0.082^* (0.030)	0.007
1942-48 vs. 1948-54	0.018 (0.010)	0.059	0.004 (0.032)	0.909
1944-50 vs. 1950-56	0.015 (0.010)	0.142	0.040 (0.034)	0.237

Table 4.6: *Estimated contrast $\hat{K}$ and $\hat{K}_o$ for breast cancer for females using 2-year age and period intervals. Corresponding p-values are given to the right of the contrast. The standard deviations of $\hat{K}$ and $\hat{K}_o$ are given in parentheses. Estimated contrasts significant on a 1% level are marked with $*$.*

Cohort period	$\hat{K}$	p	$\hat{K}_o$	p_o
1932-38 vs. 1938-44	-0.019 (0.008)	0.018	-0.060^* (0.021)	0.004
1934-40 vs. 1940-46	-0.019 (0.008)	0.015	-0.059^* (0.020)	0.004
1936-42 vs. 1942-48	-0.049^* (0.008)	0.000	-0.153^* (0.021)	0.000
1938-44 vs. 1944-50	-0.046^* (0.008)	0.000	-0.128^* (0.021)	0.000
1940-46 vs. 1946-52	-0.026^* (0.008)	0.001	-0.059^* (0.021)	0.006
1942-48 vs. 1948-54	0.011 (0.008)	0.199	0.001 (0.023)	0.975
1944-50 vs. 1950-56	0.015 (0.009)	0.104	0.037 (0.024)	0.123

Similarly as for colon cancer, the standard deviations are smaller for $\hat{K}$ compared to the standard deviations for $\hat{K}_o$ for both breast and testicular cancer. Also the standard deviations when using data for 2-year age and period intervals are smaller compared to the same values for yearly data. It might seem like $\hat{K}_o$ are significant in some cases where $\hat{K}$ is not. However the general conclusion for the results based on 1- and 2-year age and period intervals are similar.

Tables 4.5 and 4.6 for breast cancer, show a significant decrease when comparing periods before and during the occupation period in Norway. Although the tables do not indicate any increase, neither does Figure 4.2 for breast cancer, so that should not be of our concern. Thus we might have observed an actual wartime effect for breast cancer for females.

Table 4.7: *Estimated contrast $\hat{K}$ and $\hat{K}_o$ for testicular cancer using yearly data. Corresponding p-values are given to the right of the contrast. The standard deviations of $\hat{K}$ and $\hat{K}_o$ are given in parentheses. Estimated contrasts significant on a 1% level are marked with $*$.*

Cohort period	$\hat{K}$	p	$\hat{K}_o$	p_o
1932-38 vs. 1938-44	-0.078 (0.037)	0.043	-0.280 (0.123)	0.030
1934-40 vs. 1940-46	-0.027 (0.036)	0.451	-0.114 (0.119)	0.339
1936-42 vs. 1942-48	-0.070 (0.035)	0.043	0.242 (0.109)	0.026
1938-44 vs. 1944-50	0.078 (0.035)	0.024	0.117 (0.109)	0.283
1940-46 vs. 1946-52	0.050 (0.030)	0.094	0.244 (0.101)	0.015
1942-48 vs. 1948-54	0.024 (0.029)	0.409	0.123 (0.098)	0.206
1944-50 vs. 1950-56	0.027 (0.030)	0.366	-0.020 (0.096)	0.839

Table 4.8: *Estimated contrast $\hat{K}$ and $\hat{K}_o$ for testicular cancer using 2-year age and period intervals. Corresponding p-values are given to the right of the contrast. The standard deviations of $\hat{K}$ and $\hat{K}_o$ are given in parentheses. Estimated contrasts significant on a 1% level are marked with $*$.*

Cohort period	$\hat{K}$	p	$\hat{K}_o$	p_o
1932-38 vs. 1938-44	−0.037 (0.033)	0.296	−0.210 (0.085)	0.022
1934-40 vs. 1940-46	−0.015 (0.031)	0.636	0.006 (0.081)	0.946
1936-42 vs. 1942-48	0.051 (0.029)	0.084	0.209* (0.076)	0.006
1938-44 vs. 1944-50	0.066 (0.030)	0.026	0.062 (0.077)	0.424
1940-46 vs. 1946-52	0.033 (0.026)	0.203	0.163 (0.071)	0.021
1942-48 vs. 1948-54	0.025 (0.025)	0.325	0.142 (0.067)	0.034
1944-50 vs. 1950-56	0.032 (0.025)	0.211	0.018 (0.067)	0.788

For testicular cancer, the tables do not show a significant wartime effect for this particular cancer site. On the contrary, Figure 4.2 for testicular cancer indicate a wartime effect. Although the effect seems to be later than for colon cancer discussed above, we would expect the wartime effect to be captured by the estimated contrasts. The number of new cases for testicular cancer are relatively small compared to both colon and breast cancer (Cancer Registry of Norway, 2010b). Thus the standard deviations are larger which result in higher p-values. This might be, and is possibly, the explanation for the insignificant wartime effect for testicular cancer.

However from Figure 4.2 we notice that the wartime effect for testicular cancer seems to be shorter and less profound than for example for colon and breast cancer. Thus we shortened the period intervals for testicular cancer containing four years in each time interval instead of six, e.g. instead of comparing period 1932-38 to 1938-44 we compared period 1934-38 to 1938-42. For 2-year age and period intervals we obtained a significant increase in the linear trend for period 1940-44 vs. 1944-48. For yearly data the same periods where on the borderline of being significant with a p-value of 0.011. These significant increases in linear trend are more than we could accomplished with longer period intervals, so it would be wise to reconsider the length of the time intervals for testicular cancer. However as mentioned in the introduction the relatively low amount of available data in Norway might cause incomplete interpretations of the results obtained in this thesis.

To sum up the main points of this section we see that the results seem to be rather similar if we use 1- or 2-year age and period intervals. However the standard deviations are smaller for 2-year age and period intervals for all three cancer sites, which is reasonable, since the variation decreases with

the increasing number of new cases in each cell. For all three cancer sites the calculated estimated contrasts verify the trends of the estimated birth cohorts given in Figure 4.2. That does not necessarily mean that the estimated contrasts are significant. An example is testicular cancer, where the visual inspection of the estimated birth cohort effects show a wartime effect and where the estimated contrasts are insignificant. We did not see a clear trend for neither the estimated contrasts nor the corresponding p-values. We did not see a clear difference between the results for 1- or 2-year age and period intervals either. We will give similar tables for all cancer sites in Chapter 5. However, due to the smaller standard deviations, we only consider the contrast K for 2-year age and period intervals.

4.3 Nonparametric evaluation of birth cohort trends

We will try one more method in our search for a possible wartime effect for the cancer sites in Norway and consider a nonparametric method introduced by Tarone and Chu (2000). The authors suggest that the method is useful adjunct to apc models. The method is a generalization of the sign test and a simplified version of the method will be introduced in this section.

To refresh our memory the estimated rates for a given cancer site and sex are given as $\hat{r}_{apc} = d_{apc}/Y_{apc}$ for age group a, period p and birth cohort c. Further d_{apc} and Y_{apc} represent the number of new cases and person-years, as defined in section 2.1. The nonparametric method is based on the the indicator functions

$$N_{a,p,c} = \begin{cases} 1 & \text{if } \hat{r}_{a,p+1,c+1} < \hat{r}_{a,p,c} \\ 0 & \text{otherwise} \end{cases}$$

Hence $N_{a,p,c}$ will be given the value 1 if the rate for age group a in birth cohort $c+1$ is lower than in birth cohort c, i.e. lower in period $p+1$ than in period p, and 0 otherwise. Although the probability of two rates, for two different birth cohorts, being equal is very small, we will let $N_{a,p,c} = 0$ in the situations where this is the case.

As an example we consider a table of observed rates with 5 periods, 9 age groups which leads to 13 birth cohorts:

	Calendar periods compared				
Age group	1	2	3	4	5
1	$r_{1,1,9}$	$r_{1,2,10}$	$r_{1,3,11}$	$r_{1,4,12}$	$r_{1,5,13}$
2	$r_{2,1,8}$	$r_{2,2,9}$	$r_{2,3,10}$	$r_{2,4,11}$	$r_{2,5,12}$
3	$r_{3,1,7}$	$r_{3,2,8}$	$r_{3,3,9}$	$r_{3,4,10}$	$r_{3,5,11}$
4	$r_{4,1,6}$	$r_{4,2,7}$	$r_{4,3,8}$	$r_{4,4,9}$	$r_{4,5,10}$
5	$r_{5,1,5}$	$r_{5,2,6}$	$r_{5,3,7}$	$r_{5,4,8}$	$r_{5,5,9}$
6	$r_{6,1,4}$	$r_{6,2,5}$	$r_{6,3,6}$	$r_{6,4,7}$	$r_{6,5,8}$
7	$r_{7,1,3}$	$r_{7,2,4}$	$r_{7,3,5}$	$r_{7,4,6}$	$r_{7,5,7}$
8	$r_{8,1,2}$	$r_{8,2,3}$	$r_{8,3,4}$	$r_{8,4,5}$	$r_{8,5,6}$
9	$r_{9,1,1}$	$r_{9,2,2}$	$r_{9,3,3}$	$r_{9,4,4}$	$r_{9,5,5}$

This leads to the corresponding matrix with the $N_{a,p,c}$ values similar to Table 1 in Tarone and Chu (2000):

	Calendar periods compared			
Age group	2 to 1	3 to 2	4 to 3	5 to 4
1	$N_{1,1,9}$	$N_{1,2,10}$	$N_{1,3,11}$	$N_{1,4,12}$
2	$N_{2,1,8}$	$N_{2,2,9}$	$N_{2,3,10}$	$N_{2,4,11}$
3	$N_{3,1,7}$	$N_{3,2,8}$	$N_{3,3,9}$	$N_{3,4,10}$
4	$N_{4,1,6}$	$N_{4,2,7}$	$N_{4,3,8}$	$N_{4,4,9}$
5	$N_{5,1,5}$	$N_{5,2,6}$	$N_{5,3,7}$	$N_{5,4,8}$
6	$N_{6,1,4}$	$N_{6,2,5}$	$N_{6,3,6}$	$N_{6,4,7}$
7	$N_{7,1,3}$	$N_{7,2,4}$	$N_{7,3,5}$	$N_{7,4,6}$
8	$N_{8,1,2}$	$N_{8,2,3}$	$N_{8,3,4}$	$N_{8,4,5}$
9	$N_{9,1,1}$	$N_{9,2,2}$	$N_{9,3,3}$	$N_{9,4,4}$

In the latter table the comparison of two adjacent birth cohorts are given in the downward diagonal rows. However since we are interested in examining possible changes in the disease risk in birth cohorts we obtain a table similar to Table 2 in Tarone and Chu (2000), where the comparison between the adjacent birth cohorts are given on the horizontal rows. Thus if the disease rates have increased from birth cohort $c + 1$ compared to birth cohort c the corresponding row should be dominated with zeroes. Otherwise the corresponding row should be dominated with ones.

	Calendar periods compared			
Cohort compared	2 to 1	3 to 2	4 to 3	5 to 4
13 to 12				$N_{1,4,12}$
12 to 11			$N_{1,3,11}$	$N_{2,4,11}$
11 to 10		$N_{1,2,10}$	$N_{2,3,10}$	$N_{3,4,10}$
10 to 9	$N_{1,1,9}$	$N_{2,2,9}$	$N_{3,3,9}$	$N_{4,4,9}$
9 to 8	$N_{2,1,8}$	$N_{3,2,8}$	$N_{4,3,8}$	$N_{5,4,8}$
8 to 7	$N_{3,1,7}$	$N_{4,2,7}$	$N_{5,3,7}$	$N_{6,4,7}$
7 to 6	$N_{4,1,6}$	$N_{5,2,6}$	$N_{6,3,6}$	$N_{7,4,6}$
6 to 5	$N_{5,1,5}$	$N_{6,2,5}$	$N_{7,3,5}$	$N_{8,4,5}$
5 to 4	$N_{6,1,4}$	$N_{7,2,4}$	$N_{8,3,4}$	$N_{9,4,4}$
4 to 3	$N_{7,1,3}$	$N_{8,2,3}$	$N_{9,3,3}$	
3 to 2	$N_{8,1,2}$	$N_{9,2,2}$		
2 to 1	$N_{9,1,1}$			

The purpose of introducing this method is to identify a possible significant increase or decrease in the incidence rates between two birth cohorts for all ages. We start by summing up the values of $N_{a,p,c}$ for each row in the latter table where the rowsums can take values from $0, \ldots, P-1$ (remember that P is the total number of periods under consideration). Here a rowsum of 0 indicate an increase in all rates from birth cohort c compared to birth cohort $c+1$ and where a rowsum of $P-1$ represent decrease in incidence rates for all age groups from birth cohort c compared to birth cohort $c+1$. As for this example, the rowsums can take the values 0, 1, 2, 3 and 4.

Next we have to decide how low or how high a rowsum in the latter table has to be for us to consider it as an unusual value that would not occur by chance. Tarone and Chu (2000) introduce two ways of approaching the problem at hand, but we will only consider one of these.

Under the null hypothesis that incidence rates do not change between two adjacent birth cohorts the distribution of the number of decreases can be seen as the number decreases in adjacent integers in all possible permutations of the integers $1, \ldots, P$. The permutations can easily be done by using software where we will do 1000 such permutations. If the difference in rates from birth cohort c to birth cohort $c+1$ is statistically significant then the rowsum should be amongst the 0.5% most extreme values on either side of the distribution of the permuted values for a two-sided test with level 1%.

The method was tried on for both 1- and 2-year age and period intervals, which represents 2- and 4-year overlapping birth cohorts, for colon, breast and testicular cancer. For yearly data there was no significant decrease for

the birth cohorts around WWII. However for testicular cancer we found a significant decrease, on a 1% significance level, from birth cohort 1936 (1936-1937) to 1938 (1938-1939). As mentioned in the introduction and Wanderås et al. (1995), the reduction in incidence rates for testicular cancer was observed for those born during and slightly before the occupation period. Thus the significant decrease on a 1% level in the incidence rates seem somewhat reasonable.

Now for 2-year age and period intervals the results for colon, breast and testicular cancer are given in Table 4.9.

Table 4.9: Rowsum for colon, breast and testicular cancer by sex for 2-year age and period intervals

Birth cohorts	Colon Male	Colon Female	Breast	Testicular
1950 vs 1948	7*	10	9*	12
1948 vs 1946	8*	5*	14	11
1946 vs 1944	19*	10	12	5*
1944 vs 1942	9*	15	8*	14
1942 vs 1940	8*	8*	8*	14
1940 vs 1938	14	9*	13	12
1938 vs 1936	9*	12	7*	18*

The decrease in incidence rates between the two periods is considered significant if the row sum is above 17. Similarly the increase is significant if the row sum is less than 9. Thus we notice that there is a significant decrease in rates for birth cohort 1946 $(1945-1948)$ compared to birth cohort 1944 $(1943-1946)$ for colon cancer for males. For females we found no significant decrease, although it should be mentioned that the row sum for the birth cohort 1944 $(1943-1946)$ compared to 1942 $(1942-1944)$ is on the borderline of being significant. Further the only significant decrease in incidence rates is from birth cohort 1936 $(1935-1938)$ to 1938 $(1937-1940)$ for testicular cancer, where the decrease for similar birth cohorts were also obtained for yearly data as mentioned above.

Similarly as in section 4.2.3, we would like to compare the birth cohorts for those experiencing puberty before, during and after the occupation period for breast cancer. For yearly data we are not able to find any significant effects. For data with 2-year age and period intervals, we are able to find significant decrease when comparing birth cohort 1926 to 1924 and when comparing birth cohort 1932 to 1930. However as the cohort period are relatively short, we should be careful not to overinterpret the results.

If the method used in this section should be considered reliable in this study, we would expect also to find a significant decrease in the incidence rates for the birth cohorts around WWII for colon and breast cancer for females. A possible explanation for the insignificant findings might be that the incidence rates for the birth cohorts from before and during the occupation period changed gradually for these specific cancer sites and it will therefore not be apparent when comparing only two adjacent birth cohorts. Thus we should compare the observed rates for two groups of cohort periods at the time, as in Tarone and Chu (2000), instead. Unfortunately due to limited time this was not considered in this thesis.

Chapter 5

Results for all cancer sites

In the previous chapter we introduced two formal tests aid in determining if the transient reduction in incidence rates for some of the cancer sites are real or random. In this chapter, we will try the first of the two tests on all the cancer sites and discuss the results. Due to the fact that we only compare two adjacent cohorts for the non-parametric method introduced in section 4.3, we do not find the second test appropriate for the studies of this chapter. For that to be the case, we would have to extend the test to compare two groups of successive birth cohorts instead. However, due to the shortage of time this has not been considered in this thesis.

We will compare the cohort periods before, during and after the occupation period in a similar way as given in section 4.2.3. For simplicity we define a *significant decrease* as a significant negative contrast when comparing the cohort period before and during the occupation period. Similarly we will use the term *significant increase* when comparing the cohort period during and after the occupation period and obtaining a significant positive contrast. In addition we define the term *full wartime effect* when observing both a significant decrease and increase for a specific cancer site by sex. For cancer in the genital organs and breast cancer for females, we will consider those being in puberty before, during and after the occupation period as well, see section 5.2.

For our discussion we group the cancer sites in the following groups:

- digestive organs
- female genital organs and breast cancer
- male genital organs
- urinary organs
- lymphoid and haematopoietic tissue

- others.

Details of which cancer sites will be considered in each group are given in the corresponding sections. In each section a table of results will be given. However, for clarity the tables only give estimated contrasts that are significant on a 5% significance level. In addition we will include figures corresponding to those given in Appendix B. That is, the estimated effects for birth cohorts by integrating natural regression splines with eight knots and the apc model for 2-year age and period intervals. We hope to increase the strength of our interpretations by including the figures together with the results of the statistical test. From the details and discussions given in section 3.5, we know that the visual inspections of the estimated effects might change only by changing the number of knots in the spline function. Thus, the reader should be careful not to overinterpret the figures.

5.1 Digestive organs

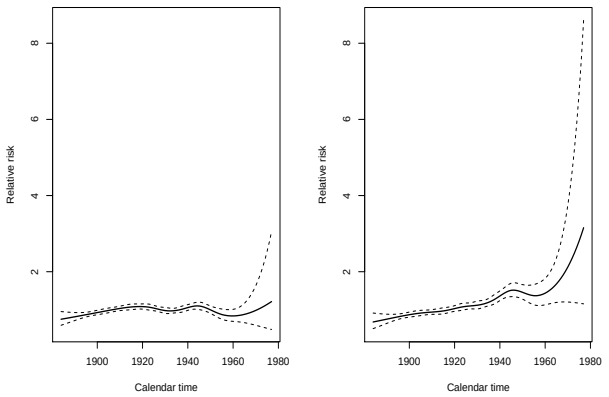
In this section we will discuss the estimated contrasts of the cancer sites which belong to digestive system, that is stomach, colon, rectum, rectosigmoid, anus and pancreas cancer. As discussed in the introduction, earlier studies have concluded with a wartime effect for colorectal cancer (see Svensson et al., 2002). Thus we hope to obtain the same conclusion here. In addition we also consider stomach and pancreas cancer. The results are given in Table 5.1 and figures 5.2 and 5.1.

In the plot for stomach cancer for both sexes we see that there is a steady decrease in the estimated effects over time. It does not seem that the occupation period has had any influence on the estimated effects, so the insignificant wartime effect for stomach cancer seems logical. As we hoped, the results indicate a full wartime effect for colon cancer for both sexes on a 1% significance level, which goes well together with the estimated effects in the plot. Further for rectal cancer for males, it does not seem to be a wartime effect at all, thus the insignificant contrasts are reasonable here as well. Although it does not look like the occupation period has had an effect on the estimated effects for rectal cancer for females, we observe a full wartime effect on a 1% significance level. This corresponds well with the CA-plot given in Appendix A. The estimated effects for pancreas cancer for males are rather constant over time and corresponds well with the insignificant estimated contrasts for this specific cancer site. As for pancreas cancer for females, we observe a full wartime effect and a significant decrease on a 5% and 1% significance level respectively. Thus, the digestive organs which are

significantly affected by the dietary habits in young age are colon cancer for both sexes and rectal and pancreas cancer for females.

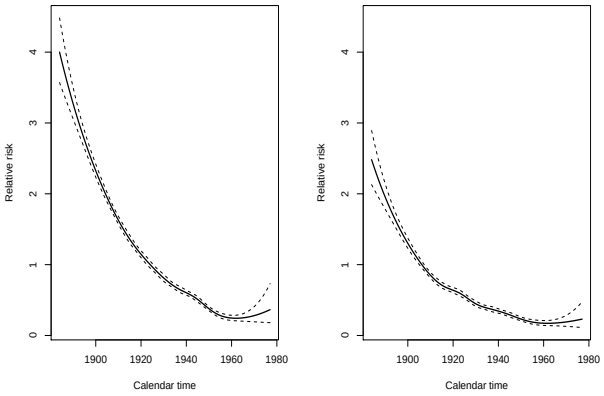
Table 5.1: *The estimated contrasts, $\hat{K}$, with the corresponding p-values. The estimated contrast significant on a 5% level are marked with $*$ and the estimated contrast significant on a 1% level are marked with $**$. The estimated standard deviations are given in parentheses.*

			Male		Female	
Site	Cohort period 1	Cohort period 2	$\hat{K}$	p	$\hat{K}$	p
Stomach	1932-38	1938-48				
	1934-40	1940-46				
	1936-42	1942-48	-0.061^{*} (0.028)	0.026		
	1938-44	1944-50				
	1940-46	1946-52				
	1942-48	1948-54				
	1944-50	1950-56				
Colon	1932-38	1938-48			-0.051^{**} (0.015)	0.001
	1934-40	1940-46			-0.046^{**} (0.017)	0.006
	1936-42	1942-48	-0.055^{**} (0.018)	0.002	-0.039^{*} (0.017)	0.023
	1938-44	1944-50	-0.056^{**} (0.018)	0.002		
	1940-46	1946-52	0.056^{**} (0.021)	0.008	0.055^{**} (0.020)	0.007
	1942-48	1948-54	0.070^{**} (0.024)	0.004	0.056^{*} (0.023)	0.013
	1944-50	1950-56				
Rectum, rectosigmoid and anus	1932-38	1938-48				
	1934-40	1940-46				
	1936-42	1942-48			-0.073^{**} (0.022)	0.001
	1938-44	1944-50				
	1940-46	1946-52				
	1942-48	1948-54			0.096^{**} (0.029)	0.001
	1944-50	1950-56				
Pancreas	1932-38	1938-48				
	1934-40	1940-46				
	1936-42	1942-48			-0.097^{**} (0.035)	0.006
	1938-44	1944-50			-0.098^{**} (0.037)	0.007
	1940-46	1946-52				
	1942-48	1948-54			0.101^{*} (0.047)	0.032
	1944-50	1950-56				

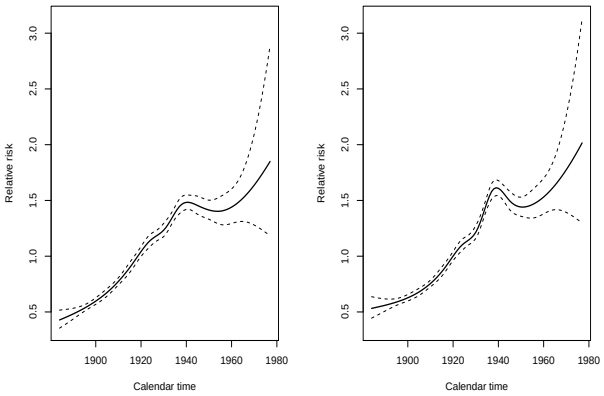


(a) Pancreas

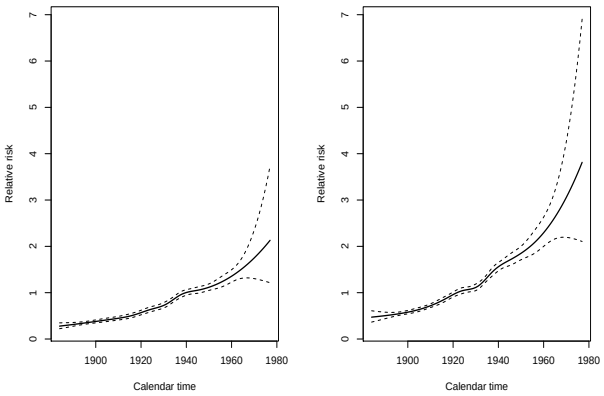
Figure 5.1: Estimated birth cohort effects for pancreas cancer in Norway 1953-2007, by using natural regression splines with eight knots. The estimated effects are obtained by using 2-year age and period intervals. The figures given on the left are for males and the figures on the right are for females.



(a) Stomach



(b) Colon



(c) Rectum, rectosigmoid and anus

Figure 5.2: Estimated birth cohort effects for stomach, colon and rectal cancer in Norway 1953-2007, by using natural regression splines with eight knots. The estimated effects are obtained by using 2-year age and period intervals. The figures given on the left are for males and the figures on the right are for females.

5.2 Female genital organs and breast cancer

The second group contains cancer in the female genital organs, i.e. cervix uteri, corpus uteri and ovary, and breast cancer for females. The estimated contrasts are given in Table 5.2 and the estimated effects in Figure 5.3.

Table 5.2: *The estimated contrasts, $\hat{K}$, with the corresponding p-values. The estimated contrast significant on a 5% level are marked with $*$ and the estimated contrast significant on a 1% level are marked with $**$. The estimated standard deviations are given in parentheses.*

Site	Cohort period 1	Cohort period 2	$\hat{K}$		p
Breast	1932-38	1938-44	-0.019^{*}	(0.008)	0.018
	1934-40	1940-46	-0.019^{*}	(0.008)	0.015
	1936-42	1942-48	-0.049^{**}	(0.008)	0.000
	1938-44	1944-50	-0.046^{**}	(0.008)	0.000
	1940-46	1946-52	-0.026^{**}	(0.008)	0.001
	1942-48	1948-50			
	1944-50	1950-56			
Cervix uteri	1932-38	1938-44			
	1934-40	1940-46			
	1936-42	1942-48			
	1938-44	1944-50			
	1940-46	1946-52			
	1942-48	1948-50	0.051^{*}	(0.020)	0.012
	1944-50	1950-56	0.041^{*}	(0.020)	0.044
Corpus uteri	1932-38	1938-44			
	1934-40	1940-46	-0.058^{**}	(0.017)	0.001
	1936-42	1942-48	-0.038^{*}	(0.017)	0.029
	1938-44	1944-50	-0.037^{*}	(0.018)	0.036
	1940-46	1946-52			
	1942-48	1948-54			
	1944-50	1950-56			
Ovary	1932-38	1938-44			
	1934-40	1940-46			
	1936-42	1942-48			
	1938-44	1944-50	-0.047^{*}	(0.019)	0.012
	1940-46	1946-52			
	1942-48	1948-54			
	1944-50	1950-56	0.046^{**}	(0.024)	0.005

For breast and corpus uteri cancer we obtain significant decrease on a 1% significance level. Similarly we found a significant increase for ovary cancer. As for cervix uteri cancer a significant increase on a 5% significance level is found. By comparing the results with the estimated effects given in Figure 5.3, the result for corpus uteri and breast cancer seem reasonable. However for cervix uteri cancer, it might be the estimated effects are subject to random variation, thus a possible wartime effect may not be easy to capture. Now, we clearly see a wartime effect by examining the plot for ovary cancer, and we obtain a full wartime effect on a 5% significance level.

Earlier studies assert that the incidence rates for breast cancer are affected for women being in puberty during the occupation period (see Tretli and Gaard, 1996). The estimated incidence rates given in Figure 5.3 imply that cervix uteri, corpus uteri and ovary cancer, might have been affected as well. Therefore, we will also consider the birth cohorts being in puberty in

the occupation period. We compared the following two groups of successive birth cohorts:

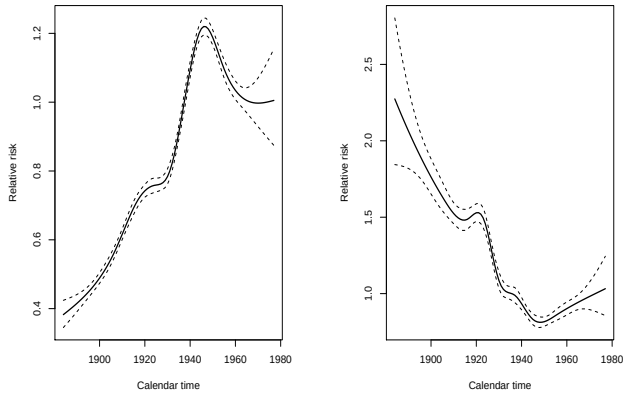
- 1920-26 vs. 1926-32
- 1922-28 vs. 1928-34
- 1924-30 vs. 1930-36
- 1926-32 vs. 1932-38
- 1928-34 vs. 1934-40.

We obtained no significant estimated contrasts for neither cervix uteri, corpus uteri nor ovary cancer on a 5% significance level. However for breast cancer, we found a significant increase in the incidence rates for the cohort period 1928-32 vs. 1932-38 on 1% significance level and 1924-30 vs. 1930-36 on 5% significance level. Although we did not find a full wartime effect for the females registered with breast cancer we believe that the incidence rates have been affected by the occupation period to some extent.

A short summary of the results obtained in this section is given here. We where not able to capture a full wartime effect for neither of the cancer sites considered in this section. We would expect to find a full wartime effect for breast cancer for females for either those experiencing puberty or for those being born around the occupation period, which we did not. However, the estimated contrast for breast, corpus uteri and ovary cancer, indicate that the dietary habits during the occupation period might have affected the birth cohorts being born during the period.

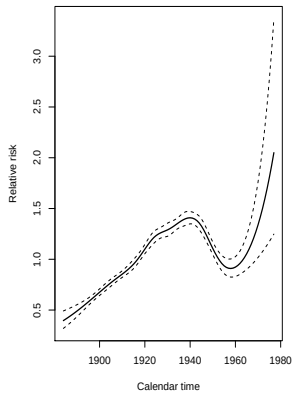
5.3 Male genital organs

In the previous section we examined the results for the genital organs for females. In this section we will consider prostate and testis cancer. Earlier studies have found that the occupation period has had an effect on the incidence rates for testicular cancer in Norway (see Wanderås et al., 1995), and the results for testicular cancer were discussed in section 4.2.3 as well. The results of the estimated contrasts for prostate and testicular cancer are given in Table 5.3 and the estimated effects for the birth cohorts are given in Figure 5.4.

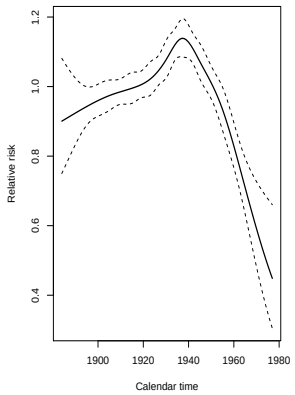


(a) Breast

(b) Cervix uteri



(c) Corpus uteri



(d) Ovary

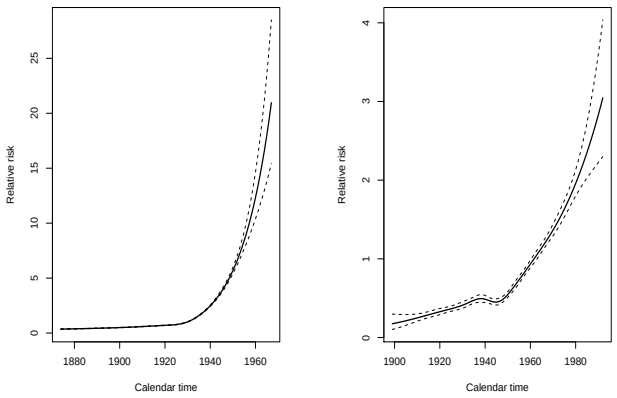
Figure 5.3: Estimated birth cohort effects for cancer in the female genital organs and breast cancer for females in Norway 1953-2007, by using natural regression splines with eight knots. The estimated effects are obtained by using 2-year age and period intervals.

For prostate cancer, the estimated effects imply a full wartime effect on a 1% significance level, which is not captured in Figure 5.4. However, by comparing the results to the CA-plot (cf. section 2.3) for prostate cancer given in Appendix A, we see that the conclusion of a full wartime effect seems resonable. Further we consider the results for testicular cancer, where we obtain no significant effects on a 1% significance level. Due to the obvious

reduction in the incidence rates which can be seen in Figure 5.4 and the fact that earlier studies indicate that there is a wartime effect for this cancer site we would hope to find an actual wartime effect for testicular cancer. However, as mentioned several times before, the relatively low amount of data available for Norway might make the results and thus the interpretations less reliable.

Table 5.3: *The estimated contrasts, $\hat{K}$, with the corresponding p-values. The estimated contrast significant on a 5% level are marked with $*$ and the estimated contrast significant on a 1% level are marked with $**$. The estimated standard deviations are given in parentheses.*

Site	Cohort period 1	Cohort period 2	$\hat{K}$	p
Prostate	1932-38	1938-44		
	1934-40	1940-46		
	1936-42	1942-48	-0.034^{**} (0.010)	0.001
	1938-44	1944-50	-0.033^{**} (0.012)	0.005
	1940-46	1946-52		
	1942-48	1948-54	0.045^{*} (0.018)	0.012
	1944-50	1950-56	0.078^{**} (0.023)	0.001
Testicular	1932-38	1938-44		
	1934-40	1940-46		
	1936-42	1942-48		
	1938-44	1944-50	0.061^{*} (0.031)	0.047
	1940-46	1946-52		
	1942-48	1948-54		
	1944-50	1950-56		



(a) Prostate (b) Testis

Figure 5.4: Estimated birth cohort effects for cancer in the male genital organs in Norway 1953-2007, by using natural regression splines with eight knots. The estimated effects are obtained by using 2-year age and period intervals.

5.4 Urinary organs

The cancer sites which are considered as urinary organs are kidney excluding renal pelvis and bladder, ureter and urethra cancer, and they will be considered in this section.

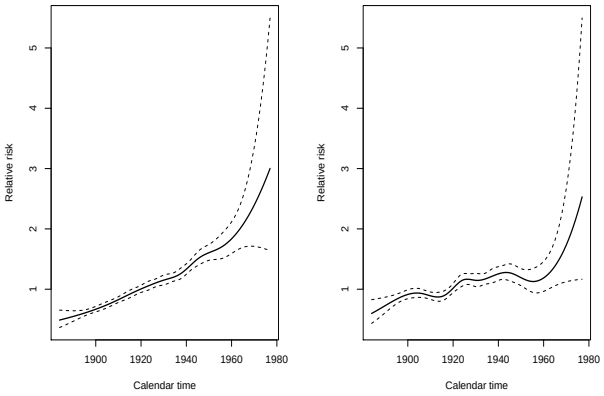
We find a full wartime effect for males experiencing kidney excluding renal pelvis cancer on a 5% significance level. Other than that, the results do not indicate that there has been a particular wartime effect for those being registered with cancer in the urinal organs. Comparing the results to Figure 5.5 the results seem reasonable. In the plots for males experiencing kidney excluding renal pelvis cancer there seems to be a small wartime effect which is captures on a 5% significance level. However, for cancer in both urinary organs for females the random variation, probably caused by relatively low number of new cases, might cause the insignificant estimated contrasts. As for bladder cancer for males, even though the plot for the estimated effects indicate that there is a transient reduction in the incidence rates, the confidence lines indicate that the estimated effects are subject to large variation in the tails.

Table 5.4: *The estimated contrasts, $\hat{K}$, with the corresponding p-values. The estimated contrast significant on a 5% level are marked with $*$ and the estimated contrast significant on a 1% level are marked with $**$. The estimated standard deviations are given in parentheses.*

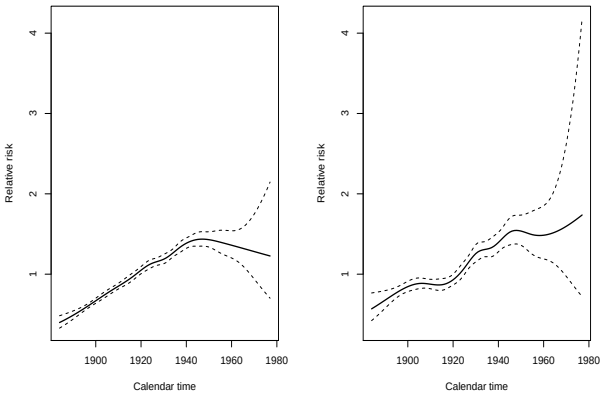
			Male			Female		
Site	Cohort period 1	Cohort period 2	$\hat{K}$		p	$\hat{K}$		p
Kidney excluding renal pelvis	1932-38	1938-44	0.054^{*}	(0.025)	0.030			
	1934-40	1940-46						
	1936-42	1942-48	-0.055^{*}	(0.025)	0.026			
	1938-44	1944-50	-0.075^{**}	(0.026)	0.004	-0.075^{*}	(0.037)	0.041
	1940-46	1946-52						
	1942-48	1948-54	0.076^{*}	(0.030)	0.011			
	1944-50	1950-54						
Bladder, ureter and urethra	1932-38	1938-44						
	1934-40	1940-46						
	1936-42	1942-48						
	1938-44	1944-50						
	1940-46	1946-52						
	1942-48	1948-54				0.103^{*}	(0.044)	0.019
	1944-50	1950-56						

5.5 Lymphoid and haematopoietic tissue

Lymphoid and haematopoietic tissue is the next group considered and is represented by non-Hodgkin lymphoma and leukemia cancer. The estimated contrasts are given in Table 5.5 and the estimated effects are given in Figure 5.6.



(a) Kidney excluding renal pelvis



(b) Bladder, ureter and urethra

Figure 5.5: Estimated birth cohort effects for cancer in the urinary organs in Norway 1953-2007, by using natural regression splines with eight knots. The estimated effects are obtained by using 2-year age and period intervals. The figures given on the left are for males and the figures on the right are for females.

As we can see from Table 5.5 there are no significant contrasts on a 1% significance level. However, we find a significant decrease for non-Hodgkin lymphoma for both sexes in addition to a significant decrease for the males experiencing leukemia cancer on a 5% significance level. Comparing the results to the plots in Figure 5.6, it might seem like there is an actual wartime effect for non-Hodgkin lymphoma cancer. In addition the confidence lines indicate that leukemia cancer might be subject to large random variation.

Table 5.5: *The estimated contrasts, $\hat{K}$, with the corresponding p-values. The estimated contrast significant on a 5% level are marked with $*$ and the estimated contrast significant on a 1% level are marked with $**$. The estimated standard deviations are given in parentheses.*

			Male			Female		
Site	Cohort period 1	Cohort period 2	$\hat{K}$		p	$\hat{K}$		p
Non-Hodgkin lymphoma	1932-38	1938-44						
	1934-40	1940-46						
	1936-42	1942-48				-0.059^*	(0.026)	0.022
	1938-44	1944-50	-0.046^*	(0.022)	0.040			
	1940-46	1946-52	-0.057^*	(0.023)	0.014			
	1942-48	1948-54	-0.059^*	(0.025)	0.018			
	1944-50	1950-56						
Leukaemia	1932-38	1938-44						
	1934-40	1940-46						
	1936-42	1942-48	-0.063^*	(0.031)	0.038			
	1938-44	1944-50						
	1940-46	1946-52						
	1942-48	1948-54						
	1944-50	1950-56						

5.6 Others

The last groups considered in this chapter are the remaining sites, i.e. mouth and pharynx, lung and trachea, melanoma of the skin, central nervous system and thyroid gland. The results are given in Table 5.6.

Here we find no significant estimated contrasts for mouth cancer for either males or females, which might be caused by sparse data. As for lung cancer, we find a significant decrease for both males and females on a 1% significance level. However, if we increase the significance level to 5% we obtain a full wartime effect for males experiencing lung cancer. The results seem reasonable when comparing the results to plots of the estimated effects for lung cancer. Further we find a full wartime effect on 1% significance level for melanoma of the skin and thyroid gland cancer, both for males. If we increase the level to 5% we find a full wartime effect for melanoma of the skin and the central nervous system cancer, both for females, as well. However, compared to Figure 5.8, we would probably expect a full wartime effect for males experiencing cancer in the central nervous system and for females experiencing cancer in the thyroid gland. Although, the CA-plot for thyroid gland cancer for females, indicate that they might be subject to large random variation.

5.7 Conclusion

We will now give a short summary of the results of the previous sections. However, in order to concentrate on the main findings we will here only

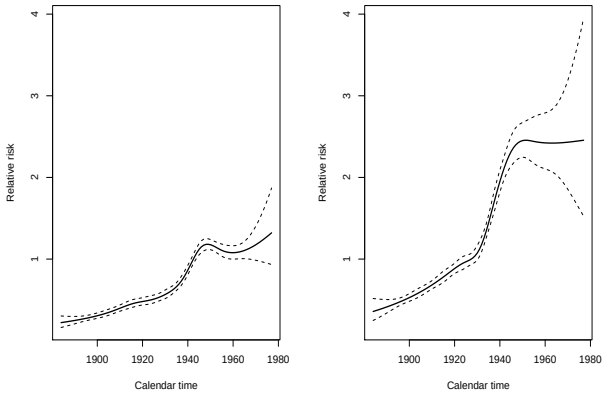
discuss the results that are significant on a 1% level. The study of this thesis was motivated with the fact that earlier studies have found a possible wartime effect for colorectal and testicular cancer and breast cancer for females in Norway. Therefore, the results for these cancer sites are of particular interest. For colon cancer we found a full wartime effect for both sexes. This indicates that dietary habits at young age are important for cancer risk later in life for this particular cancer site. Further we found a full wartime effect for rectal cancer for females as well. Thus, the study done by Svensson et al. (2002) and the results given here agree at some level. However, there were no other cancer sites which belong to the digestive system that seem to have been affected by the occupation period, except perhaps pancreas cancer for females.

As for breast cancer, Tretli and Gaard (1996) found that the females being in puberty during the occupation period were affected. We did not find a full wartime effect for breast cancer, although we did find that the incidence rates increased for those being in puberty after the occupation period compared to those being in puberty during the war. In addition we found a transient reduction in the incidence rates for the birth cohorts being born during the occupation period. Thus it seem that the dietary habits for young females and females being in puberty do affect the risk of getting breast cancer later in life. It might also seem that the dietary habits for young females might affect the risk of getting corpus uteri and ovary cancer as well.

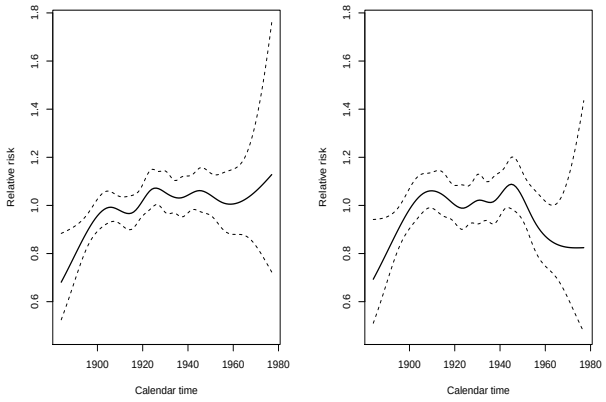
Further, we found no indication that the occupation period had an impact on the risk of testicular cancer. We think that the insignificant finding might be caused by low amount of data, without being able to say this with certainty. However, we found a full wartime effect for prostate cancer.

For the urinary organs it might seem the dietary habits for young males might affect the risk of getting kidney, excluding renal pelvis, cancer later in life. On the contrary, it does not seem to affect the risk of the cancer sites belonging to the lymphoid and haematopoietic tissue at all. However, the results indicate that the risk of getting lung cancer and melanoma of the skin for both sexes in addition to central nervous system cancer for females and thyroid gland cancer for males is affected by the dietary habits in young age.

As we can see, it seems like the dietary habits at young age affect the risk of experiencing cancer later in life for several cancer sites. In addition, the dietary habits for females in puberty might affect the risk of developing breast cancer. These results are exploratory, however further studies are needed in order to make firm conclusions, see section 6.2.



(a) Non-Hodgkin lymphoma

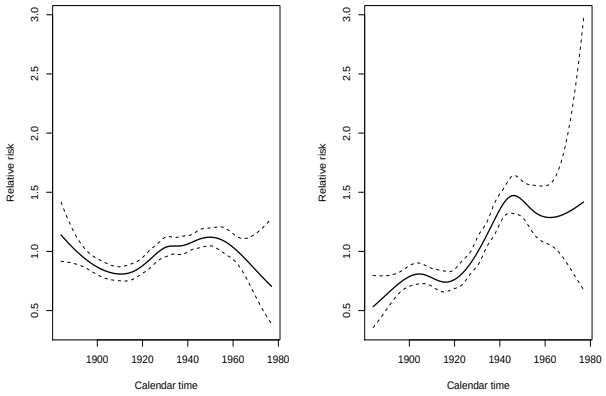


(b) Leukaemia

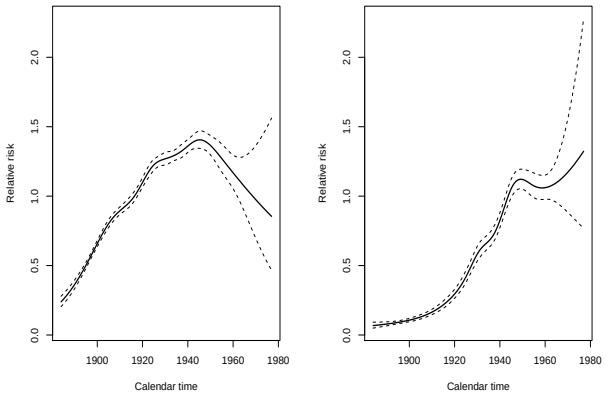
Figure 5.6: Estimated birth cohort effects for cancer in the lymphoid and haematopoietic tissue in Norway 1953-2007, by using natural regression splines with eight knots. The estimated effects are obtained by using 2-year age and period intervals. The figures given on the left are for males and the figures on the right are for females.

Table 5.6: *The estimated contrasts, $\hat{K}$, with the corresponding p-values. The estimated contrast significant on a 5% level are marked with $*$ and the estimated contrast significant on a 1% level are marked with $**$. The estimated standard deviations are given in parentheses.*

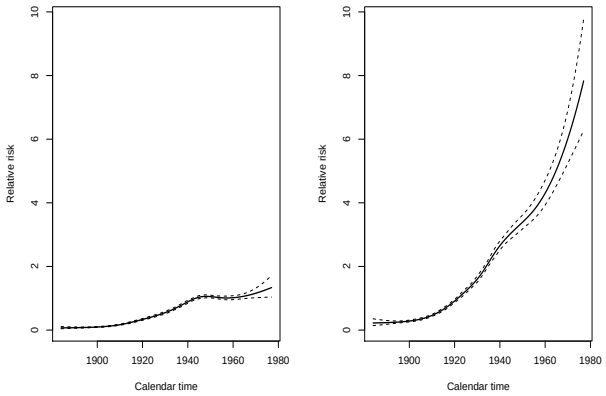
			Male		Female	
Site	Cohort period 1	Cohort period 2	$\hat{K}$	p	$\hat{K}$	p
Mouth and pharynx	1932-38	1938-44				
	1934-40	1940-46				
	1936-42	1942-48				
	1938-44	1944-50				
	1940-46	1946-52				
	1942-48	1948-54				
	1944-50	1950-56				
Lung and trachea	1932-38	1938-44				
	1934-40	1940-46				
	1936-42	1942-48	-0.030^{*} (0.014)	0.028		
	1938-44	1944-50	-0.054^{**} (0.015)	0.000	-0.063^{**} (0.017)	0.000
	1940-46	1946-52			-0.060^{**} (0.018)	0.001
	1942-48	1948-54	0.041^{*} (0.019)	0.030		
	1944-50	1950-56				
Melanoma of the skin	1932-38	1938-44				
	1934-40	1940-46			-0.038^{*} (0.017)	0.028
	1936-42	1942-48	-0.051^{**} (0.017)	0.003	-0.052^{**} (0.017)	0.002
	1938-44	1944-50	-0.081^{**} (0.017)	0.000	-0.047^{**} (0.017)	0.005
	1940-46	1946-52	-0.055^{**} (0.018)	0.002		
	1942-48	1948-54				
	1944-50	1950-56	0.058^{**} (0.021)	0.006	0.043^{*} (0.019)	0.022
Central nervous system	1932-38	1938-44				
	1934-40	1940-46				
	1936-42	1942-48			-0.073^{**} (0.021)	0.000
	1938-44	1944-50			-0.044^{*} (0.020)	0.028
	1940-46	1946-52				
	1942-48	1948-54			0.045^{*} (0.023)	0.048
	1944-50	1950-56				
Thyroid gland	1932-38	1938-44			0.081^{*} (0.033)	0.015
	1934-40	1940-46				
	1936-42	1942-48			-0.088^{**} (0.032)	0.006
	1938-44	1944-50	-0.183^{**} (0.051)	0.000		
	1940-46	1946-52	-0.112^{*} (0.052)	0.029		
	1942-48	1948-54				
	1944-50	1950-56	0.178^{**} (0.062)	0.004		



(a) Mouth and pharynx

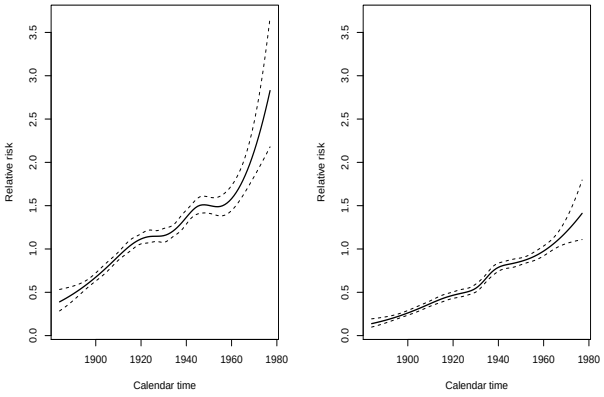


(b) Lung and trachea

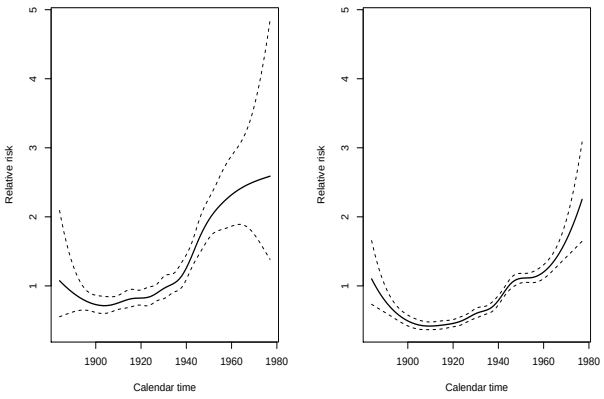


(c) Melanoma of the skin

Figure 5.7: Estimated birth cohort effects for cancer in the mouth and pharynx, lung and trachea and melanoma of the skin in Norway 1953-2007, by using natural regression splines with eight knots. The estimated effects are obtained by using 2-year age and period intervals. The figures given on the left are for males and the figures on the right are for females.



(a) Central nervous system



(b) Thyroid gland

Figure 5.8: Estimated birth cohort effects for cancer in the central nervous system and thyroid gland in Norway 1953-2007, by using natural regression splines with eight knots. The estimated effects are obtained by using 2-year age and period intervals. The figures given on the left are for males and the figures on the right are for females.

Chapter 6

Discussion

6.1 Summary

In the last chapter we sum up our main results and conclusions. Throughout this thesis we have introduced several different types of visual inspections and statistical tests as an attempt to find out if the assumed wartime effect can be considered real or random. The fact that age, period and cohort are hopelessly entangled makes the interpretations more difficult and we have to consider the so called apc model to adjust for all three variables. Fortunately, earlier studies have found appropriate ways to deal with the problem of linear dependency. That is for example extracting the drift by the method of Holford (1991), the use of natural regression splines with the apc model introduced by Heuer (1997) and the formal tests introduced by Tarone and Chu (1996, 2000). This have enhanced interpretations both visually and numerically. The figures given in Appendix B visually show the trends of the estimated effects by integrating natural regression splines and the apc model using 2-year age and period intervals. Although we have to be careful not to overinterpret the plots, they still give an overall summary of trend in incidence rates in the time before, during and after the occupation period for the different cancer sites by sex. In addition, we obtained numerical results by using a formal test in the previous chapter. The figures and numerical results are used adjunct to each other for best possible interpretations. The results indicate that the occupation period has had an impact on the incidence rates for several cancer sites. As the topic suggests, the main focus of this thesis has been to examine the impact of World War II on the cancer rates in Norway and we are somehow glad that the results support our conjecture of a wartime effect on the incidence rates

for some cancer sites.

However the number of false positives, i.e. rejections of null hypotheses that are actually true, will increase by multiple testing. As an attempt to decrease the probability of false positives we could have considered a Bonferroni correction. We perform a total of seven tests for 19 cancer sites, where 13 were performed for both sexes (we have not considered the tests made in addition for the female genital organs and breast cancer for females). Thus if we consider a Bonferroni correction we should adopt a significance level of $\frac{5\%}{224} = 0.022\%$ for each test. Then we would find a significant decrease for lung cancer and melanoma of the skin cancer for males and breast cancer for females. As expected, the indication of a possible wartime effect decreases with the use of a Bonferroni correction. However, the results given in this thesis are exploratory and as an attempt to obtain as much information on the wartime effect as possible, we have chosen not to use a Bonferroni correction in the previous chapter.

Throughout this thesis we have learned that the relatively low number of registered new cases for the cancer sites in Norway may lead to somewhat incomplete interpretations and results. Therefore we are careful not to overinterpret the findings in the previous chapter. At the same time the results keeps us motivated to study the incidence rates for the birth cohorts born during the occupation period in Norway further.

6.2 Further research

Even though the results indicate a wartime effect for some cancer sites, this should be examined more carefully. In addition we could have extended the statistical tests by controlling the false discovery rate, the expected proportion of false positives, as well (Yoav and Hochberg, 1995). Due to time limitations, we have not been able to consider this in our studies. However, as we suspect the relatively low amount of registered cancer cases to cause difficulties in our study, it might not even be worth the effort to use only Norwegian data in a possible further research. As mentioned earlier, it might be an idea to combine similar data for Denmark with the data for Norway used here. The two countries have somewhat similar trends in cancer rates for several cancer sites. An other idea would be to examine the data for Denmark alone. If the results are similar as the ones obtained in this study we might get stronger indications that the occupation period did affect the cancer rates in Norway as well as in Denmark.

Besides including data for Denmark in our studies we should expand

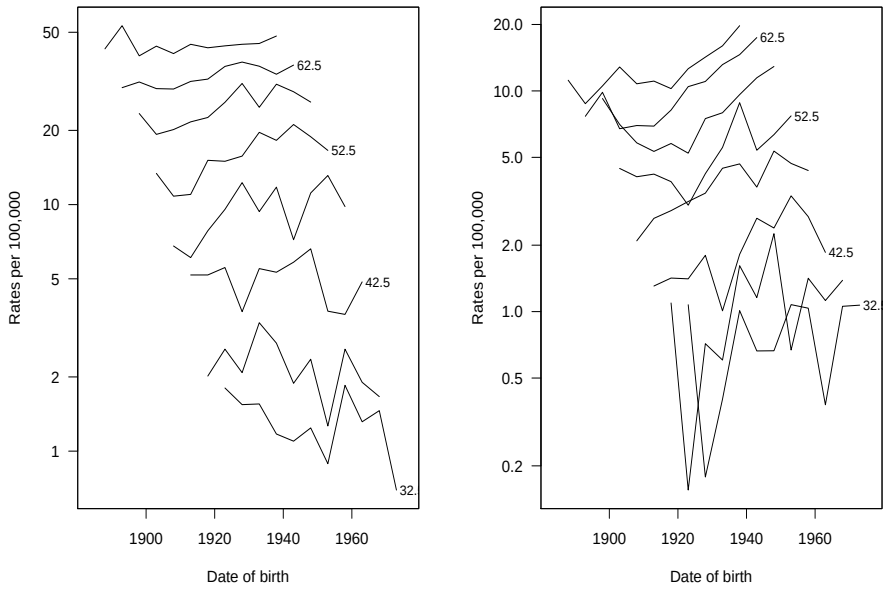
the non-parametric test in section 4.3. We think this test can be useful and provide us with reasonable results. That is, if we extend the test to compare two groups of successive birth cohorts, as in Tarone and Chu (2000), instead of only two single birth cohorts.

Thus, we conclude this study with the fact that research imply that the occupation period has had an impact on the cancer rates for Norway. However, to obtain more reliable conclusion further research is needed.

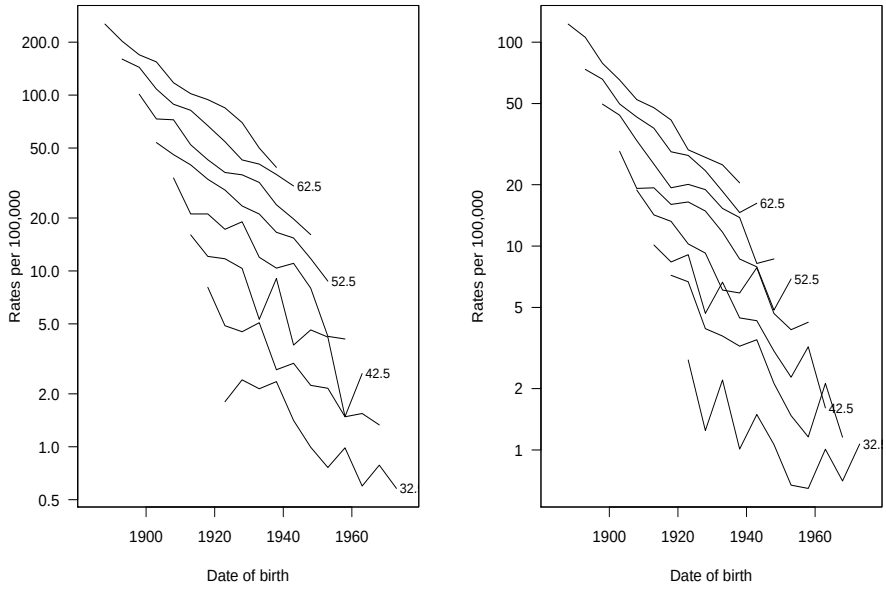
Appendix A

CA-plots for all sites

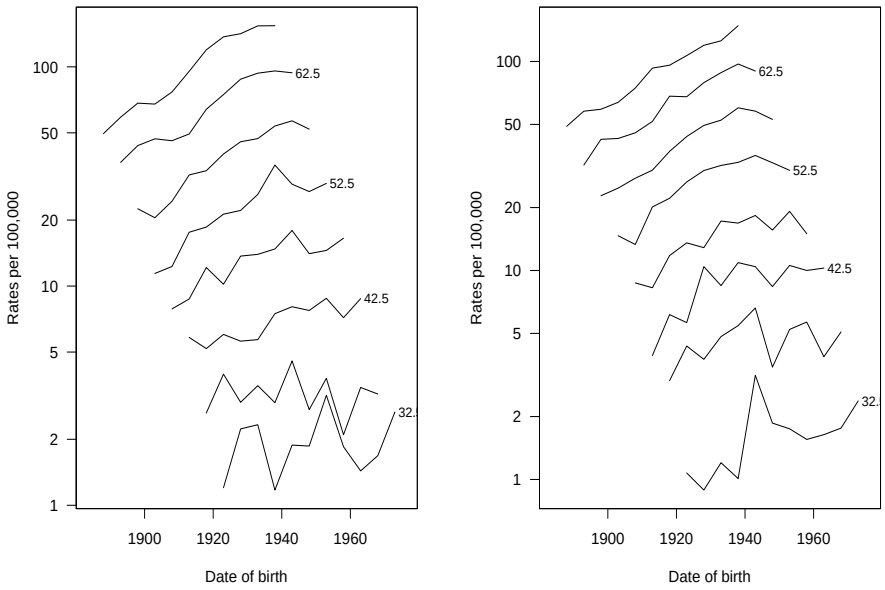
CA-plots (rates vs. cohort by age) for all cancer sites considered in this thesis. The figures given on the left are for males and the figures on the right are for females. For each age group the line represents the incidence rates over time. A possible cohort effect can be found by following a vertical line through a specific birth cohort. For further details, see section 2.3.



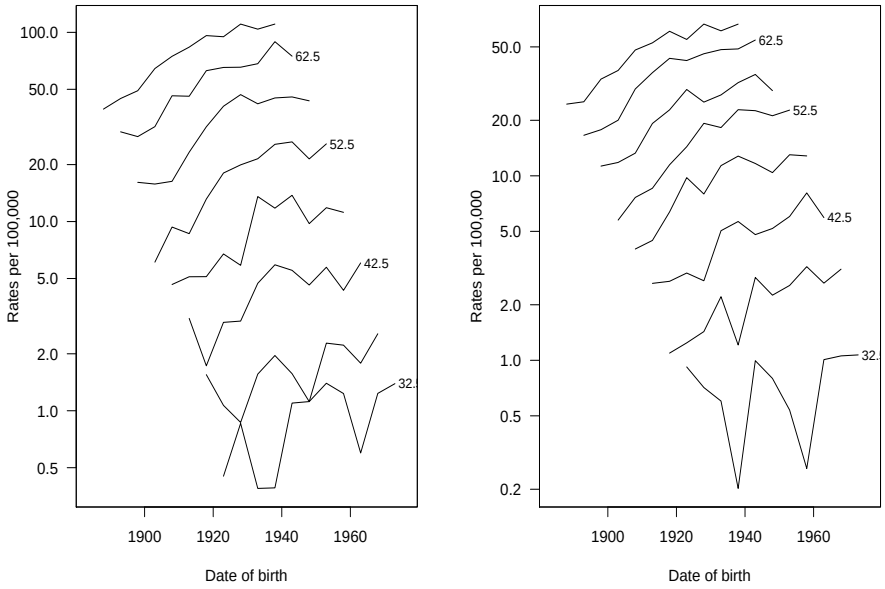
(a) Mouth and pharynx



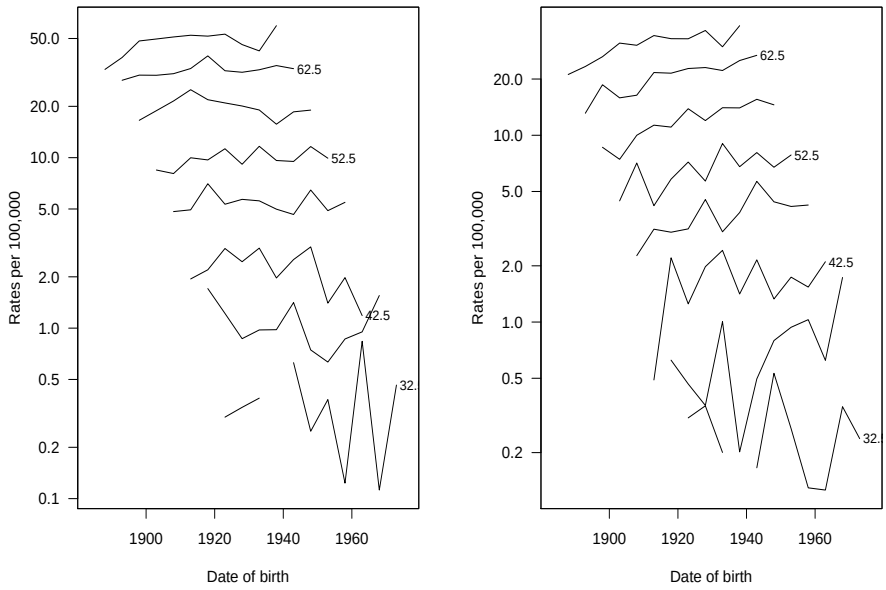
(b) Stomach



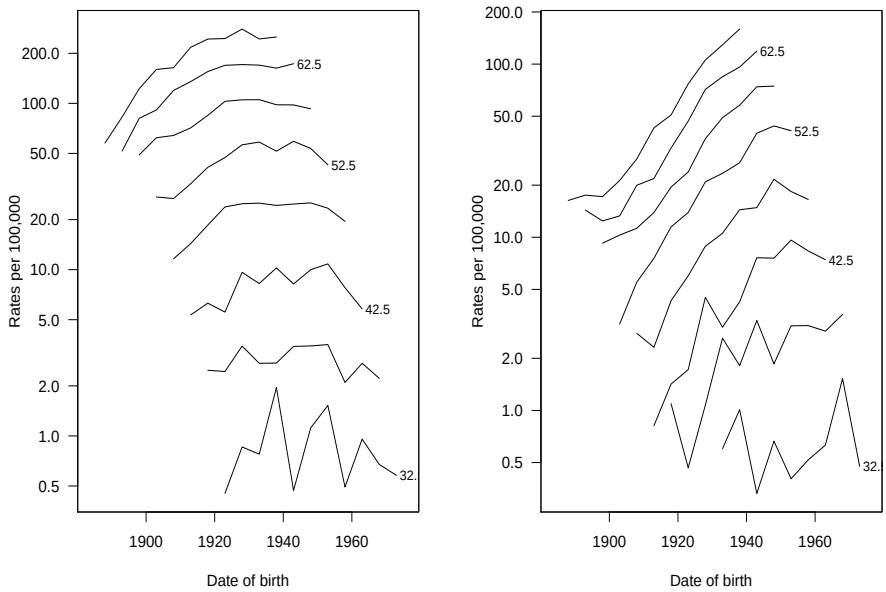
(c) Colon



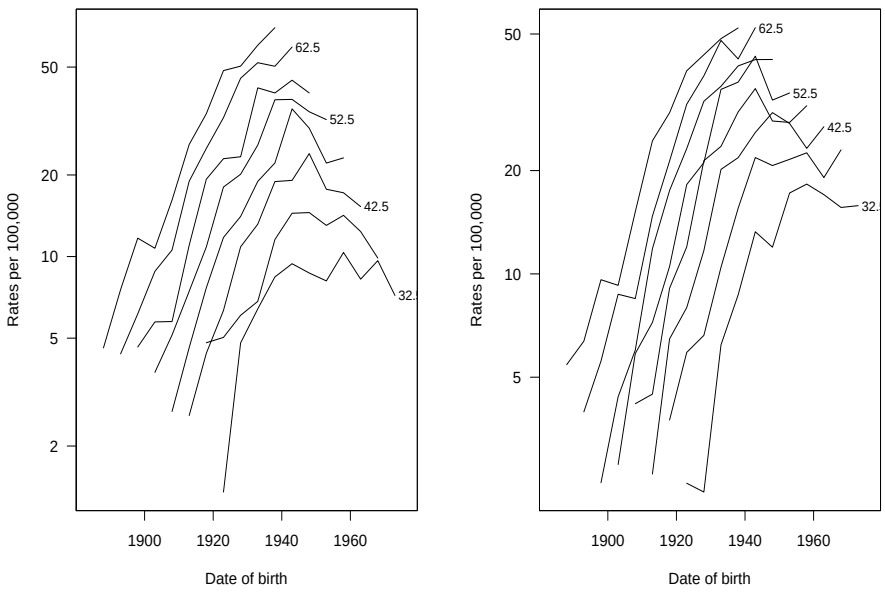
(d) Rectum, rectosigmoid and anus



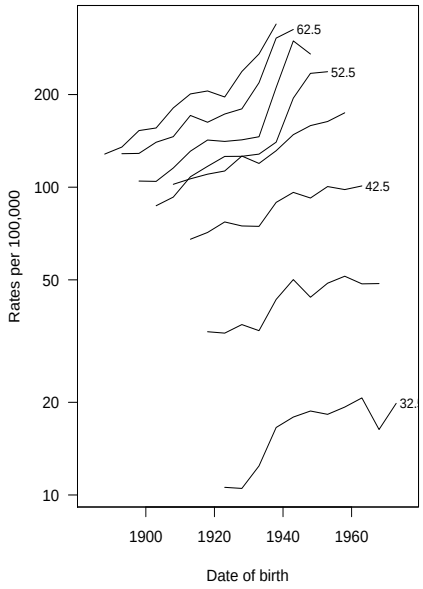
(e) Pancreas



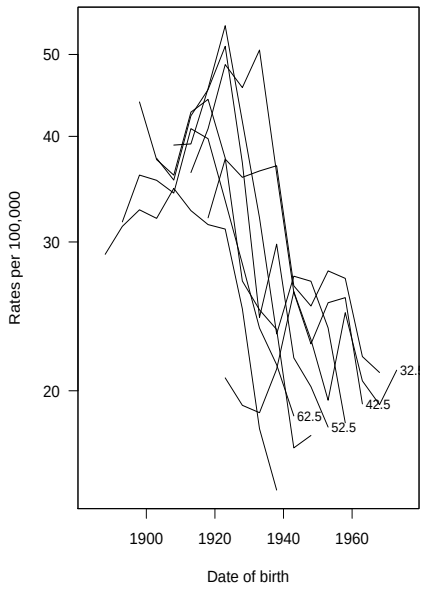
(f) Lung and trachea



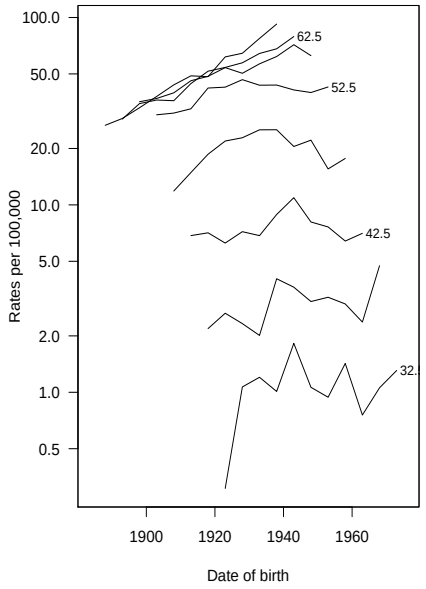
(g) Melanoma of the skin



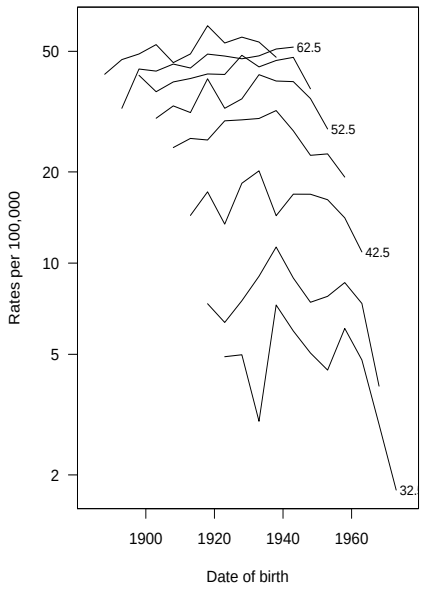
(h) Breast



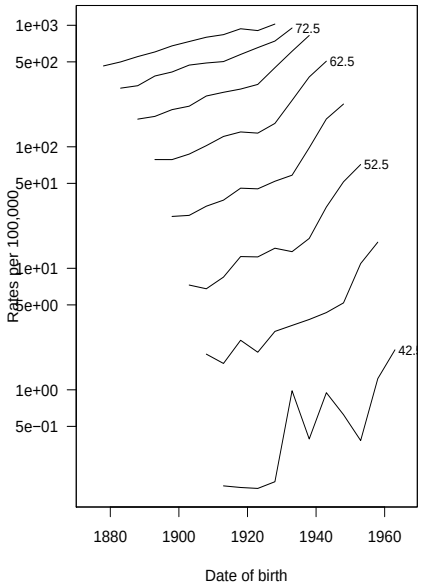
(i) Cervix uteri



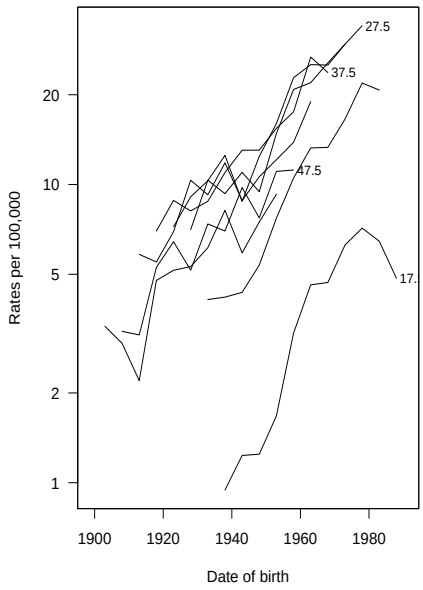
(j) Corpus uteri



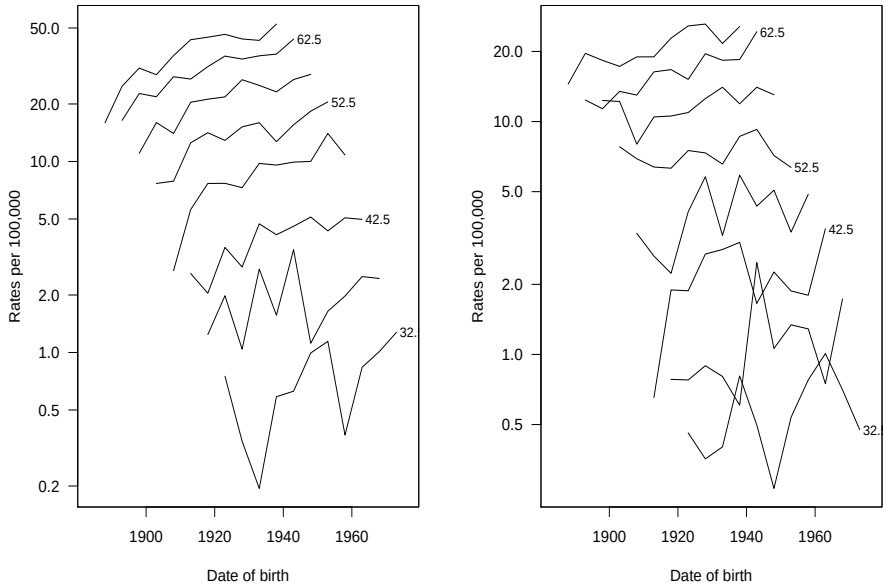
(k) Ovary



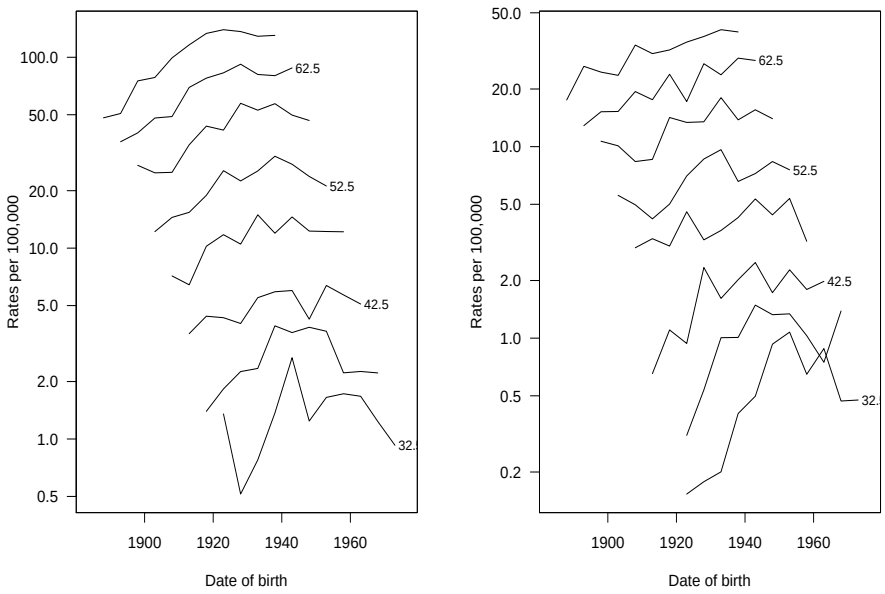
(l) Prostate



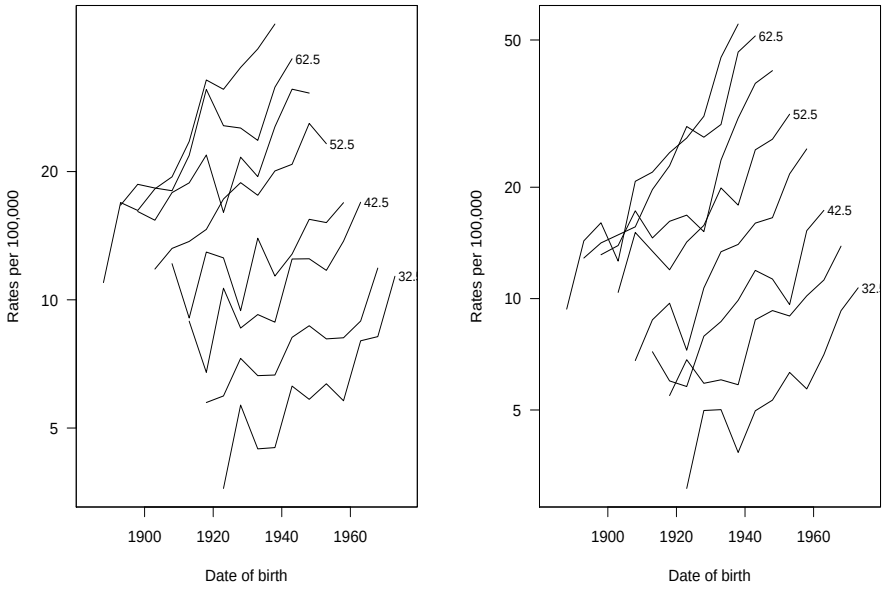
(m) Testis



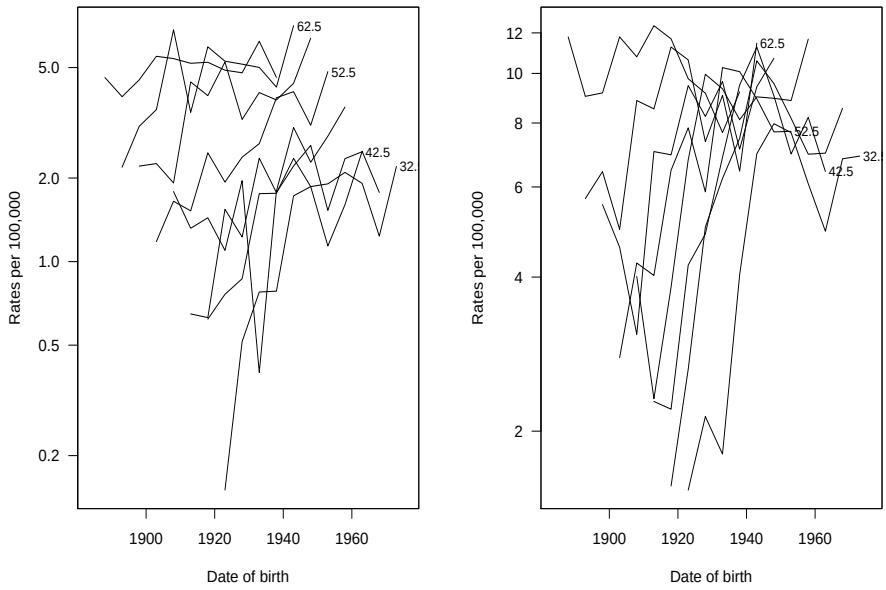
(n) Kidney excluding renal pelvis



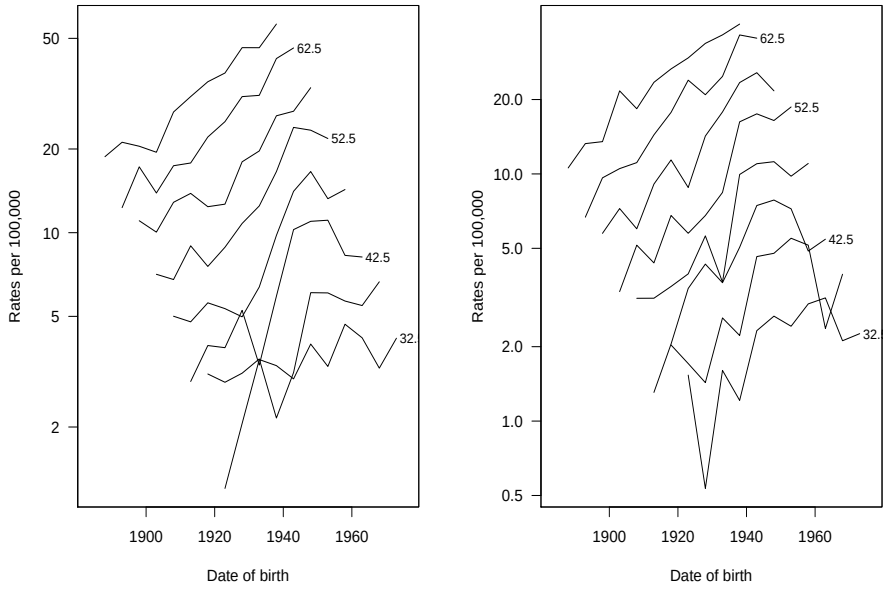
(o) Bladder, ureter and urethra



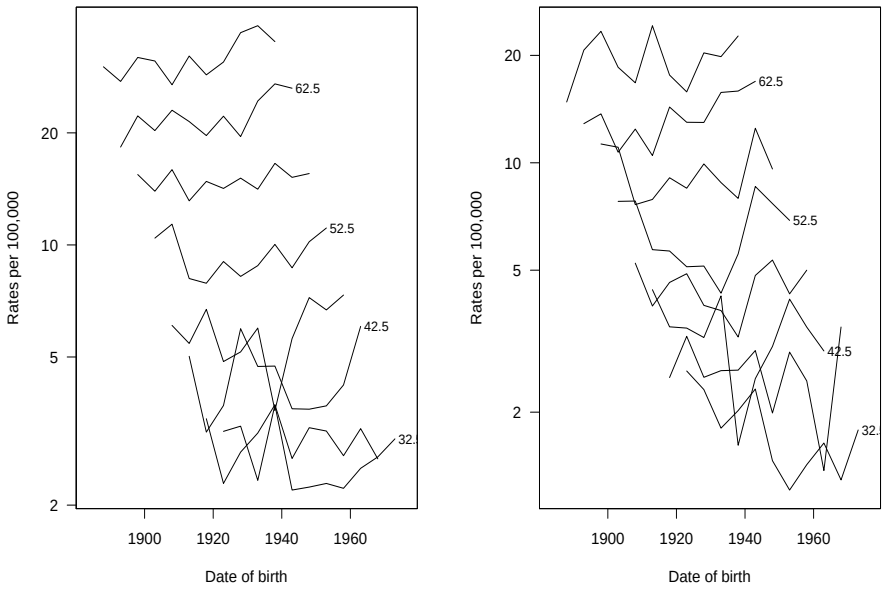
(p) Central nervous system



(q) Thyroid gland



(r) Non-Hodgkin lymphoma



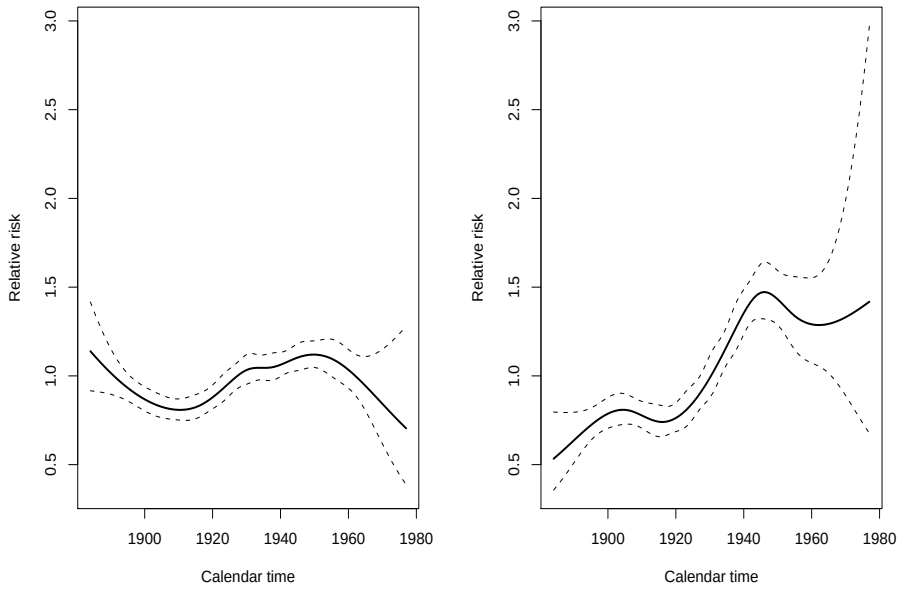
(s) Leukaemia

Figure A.1

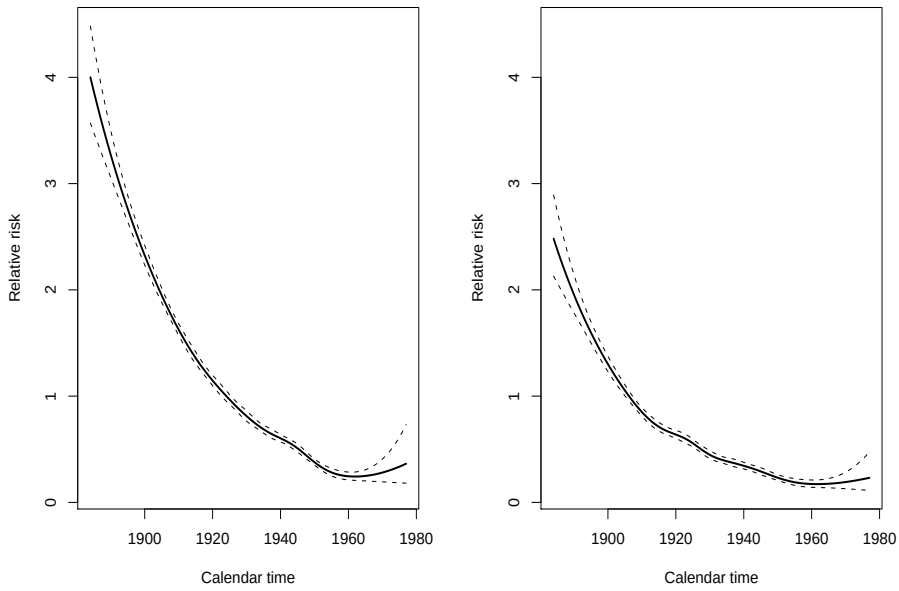
Appendix B

Estimated cohort effects using splines for all sites

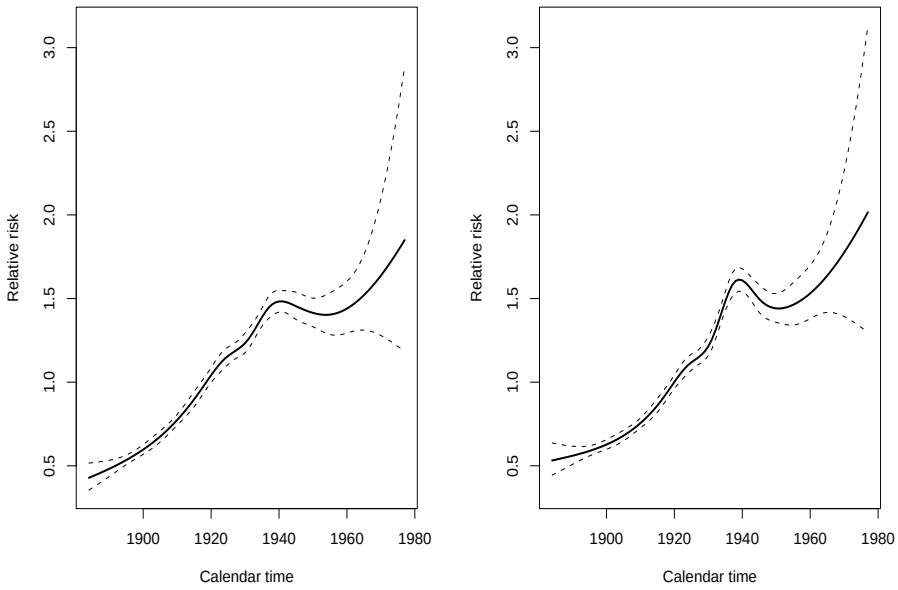
Estimated cohort effects for all cancer sites considered in this thesis. The figures given on the left are for males and the figures on the right are for females. The estimated effects are obtained by integrating natural regression splines with the apc model for 2-year age and period intervals. For further details, see section 3.5.



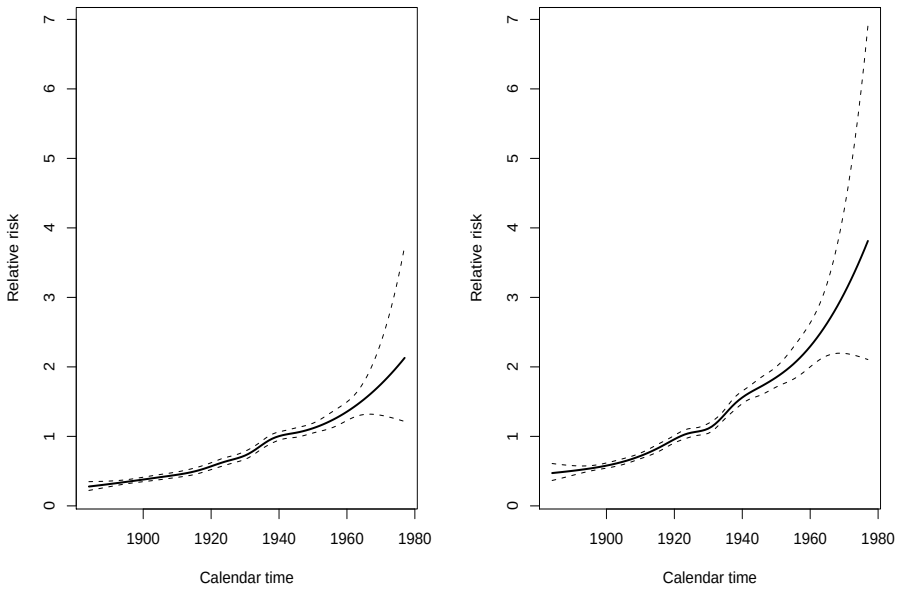
(a) Mouth and pharynx



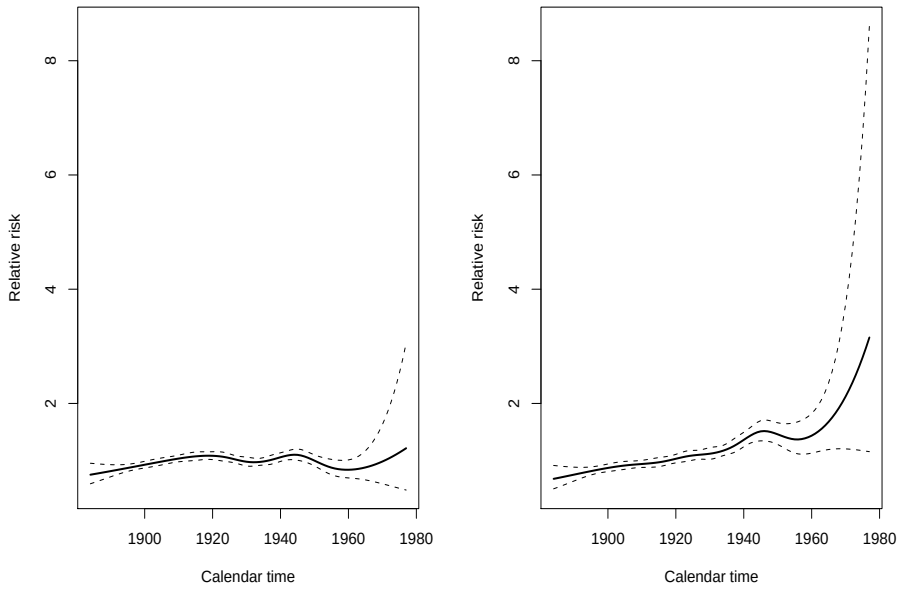
(b) Stomach



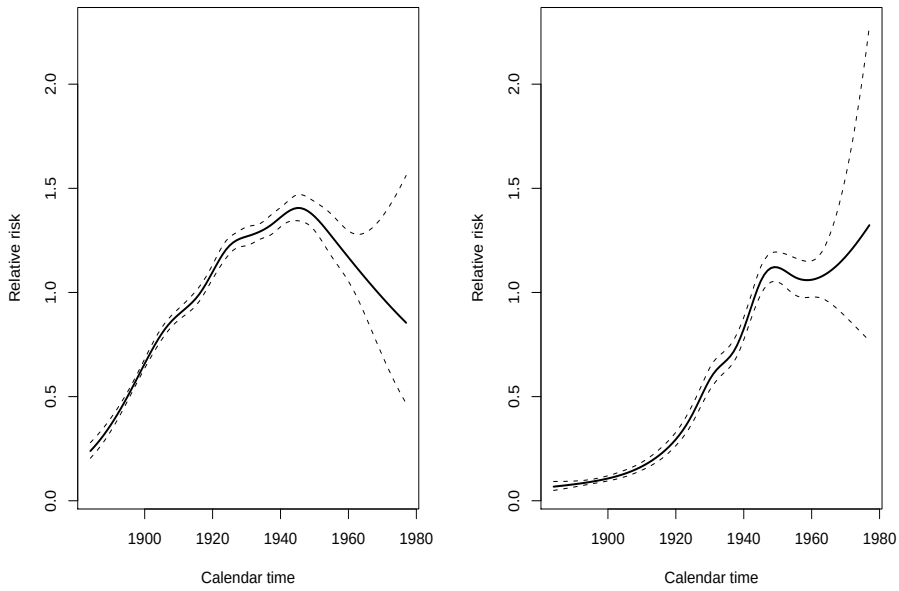
(c) Colon



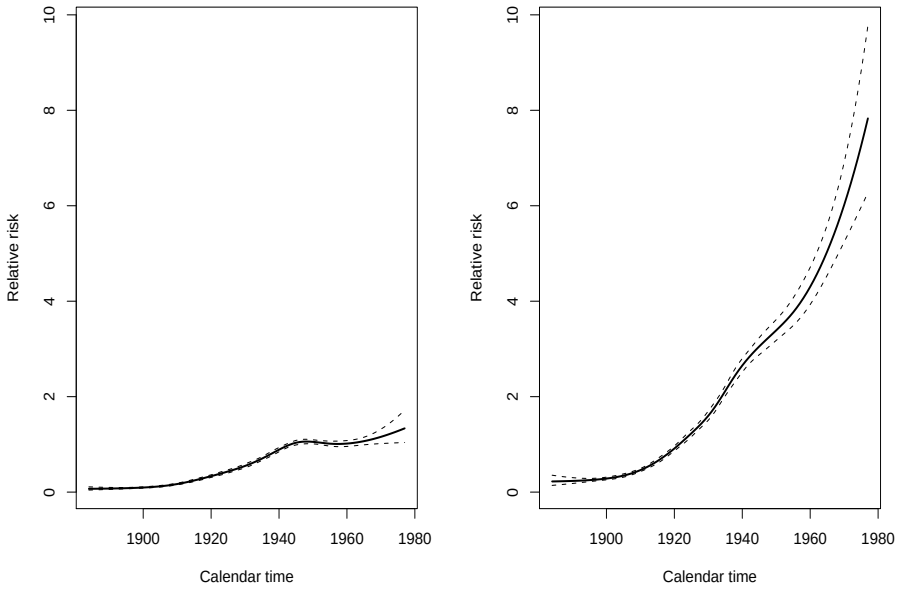
(d) Rectum, rectosigmoid and anus



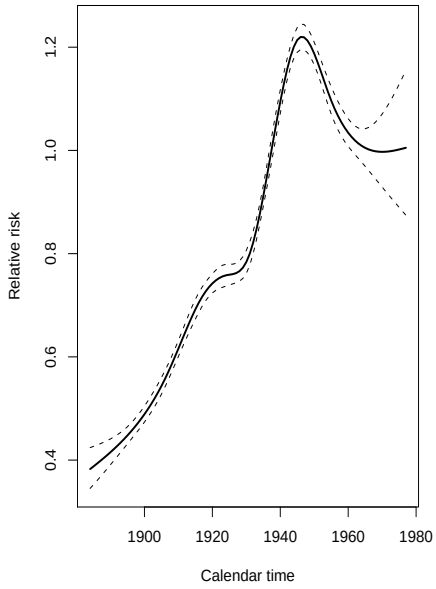
(e) Pancreas



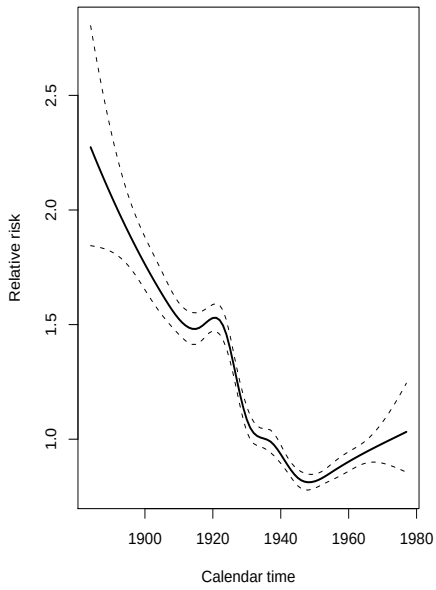
(f) Lung and trachea



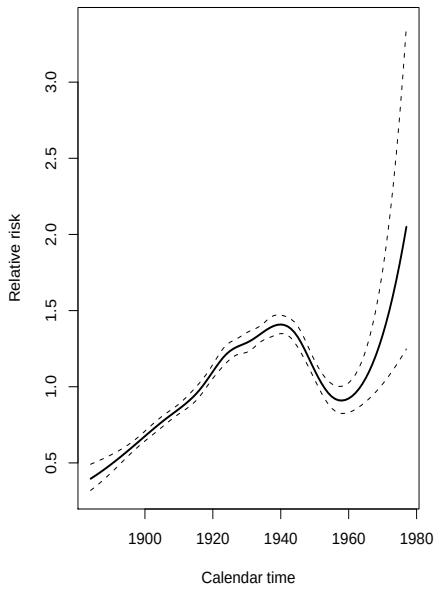
(g) Melanoma of the skin



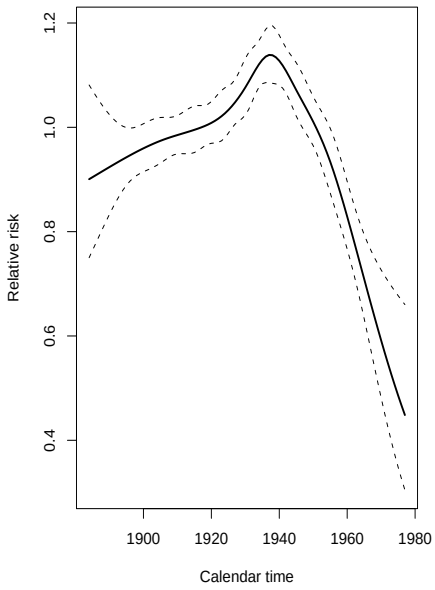
(h) Breast



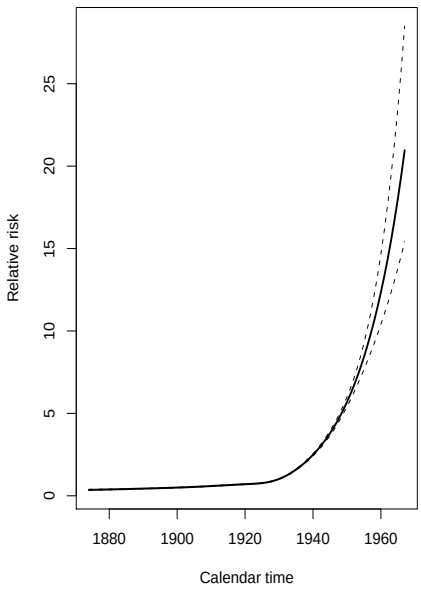
(i) Cervix uteri



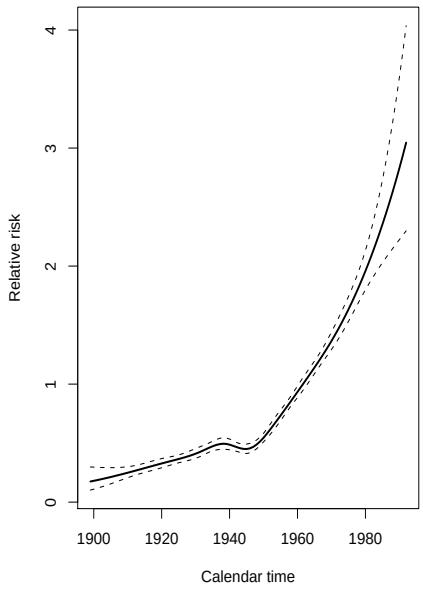
(j) Corpus uteri



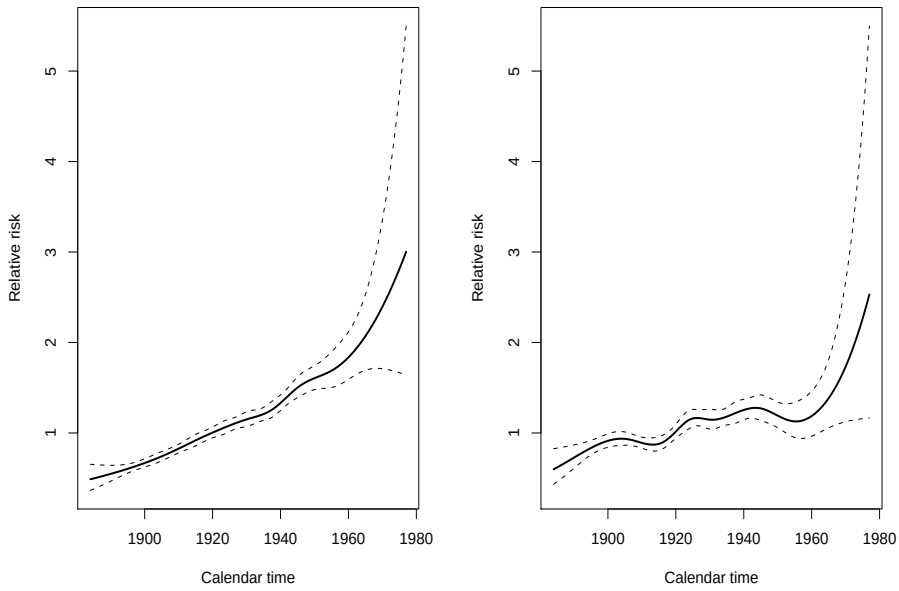
(k) Ovary



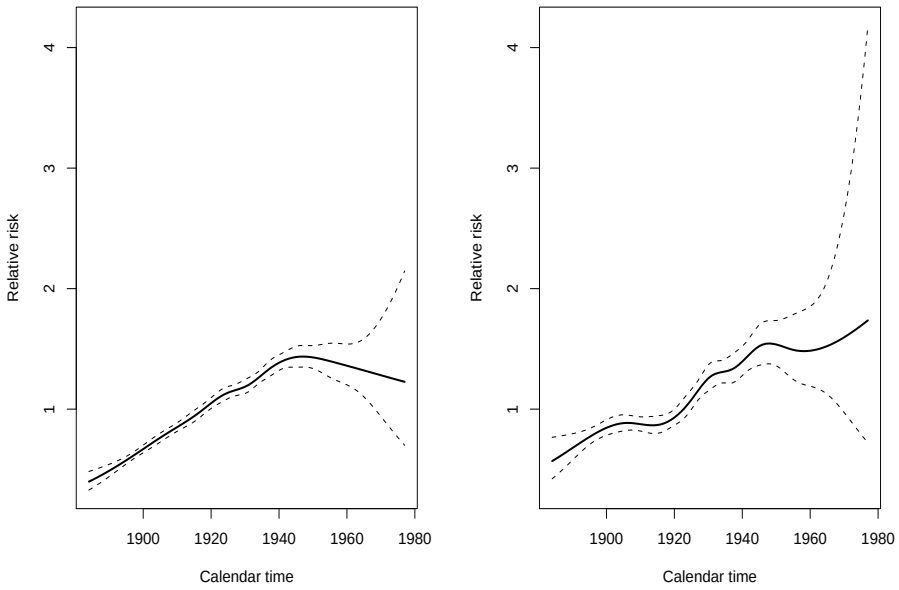
(l) Prostate



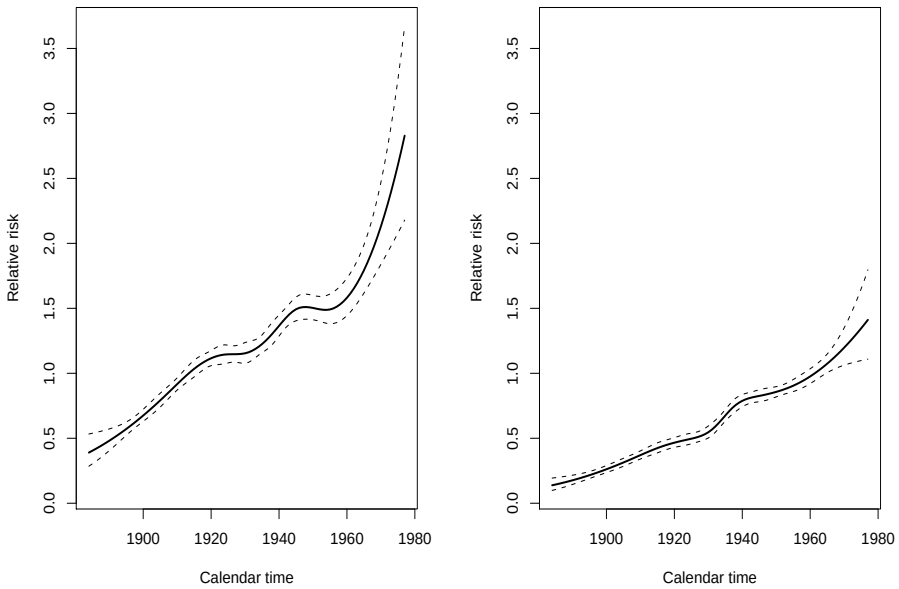
(m) Testis



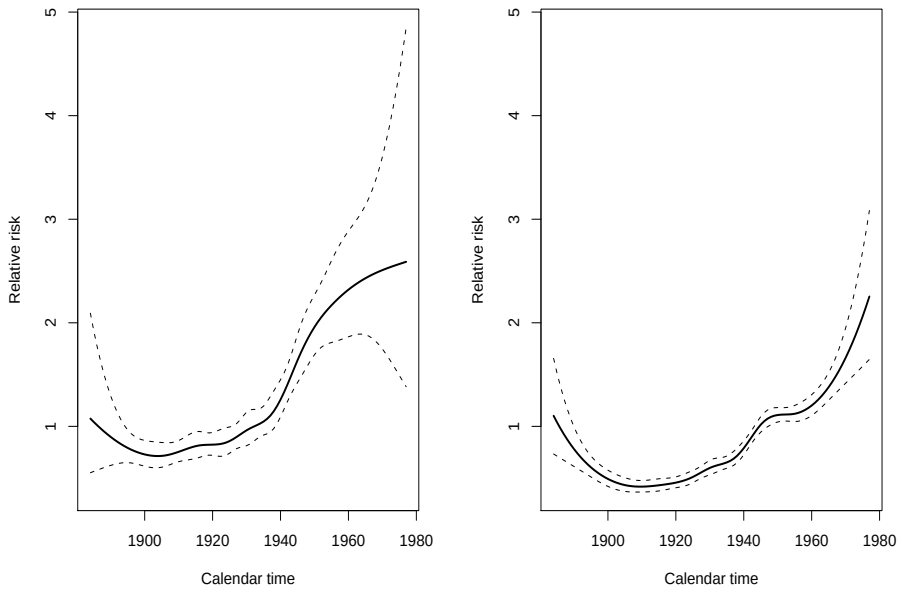
(n) Kidney excluding renal pelvis



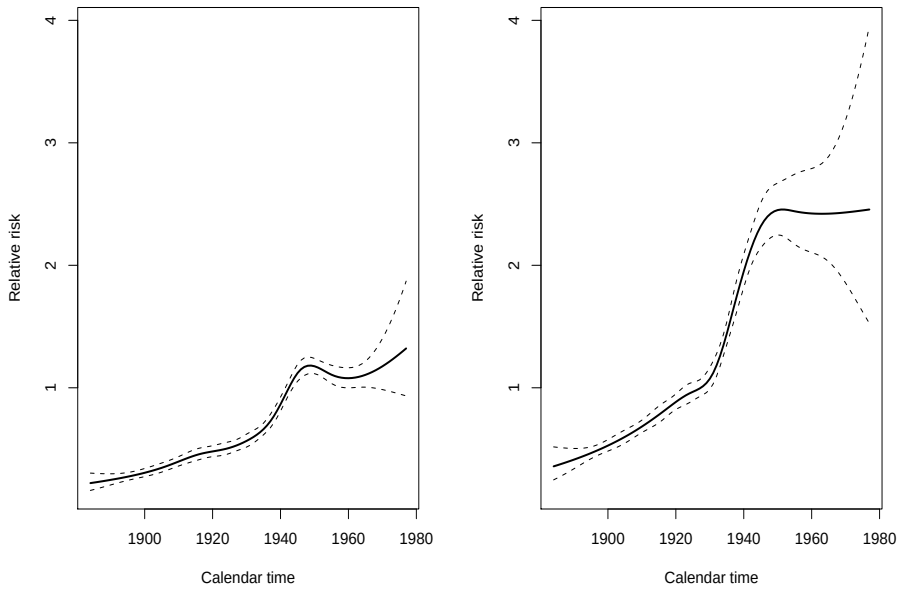
(o) Bladder, ureter and urethra



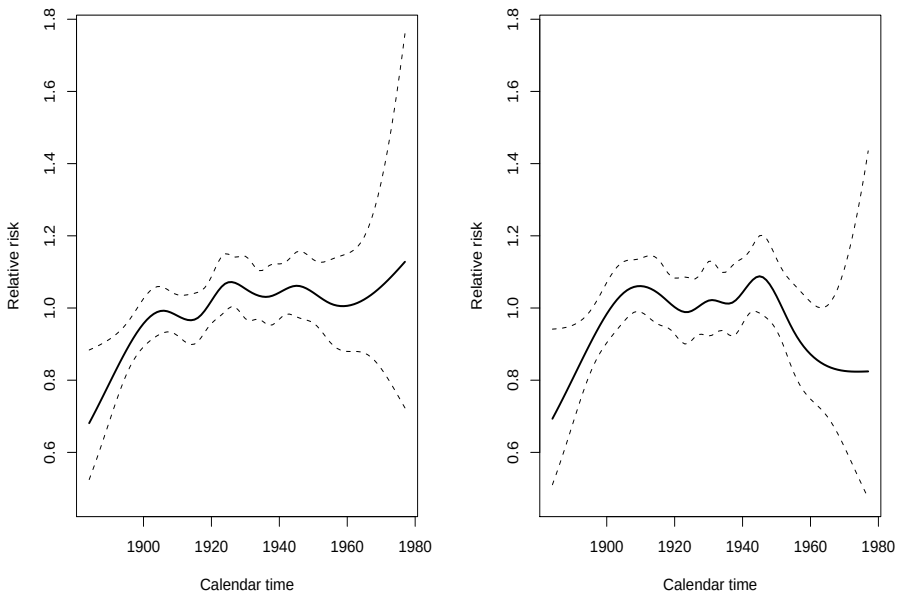
(p) Central nervous system



(q) Thyroid gland



(r) Non-Hodgkin lymphoma



(s) Leukaemia

Figure B.1

Appendix C

Orthogonal polynomials

The use of orthogonal polynomials in section 4.2.2 needs to be properly introduced. It is not necessary to read this appendix for the understanding of the tables and conclusions given in section 4.2.2 and therefore it can be skipped if the details are not of interest for the reader.

We start with a general introduction of the Gram-Schmidt process (e.g. Lay, 2006), i.e. define $\gamma_j = \theta x_{0j} + \eta_1 x_{1j} + \ldots + \eta_p x_{pj}$. Now replace the set of vectors $\{x_0, x_1, \ldots, x_p\}$ with an orthogonal set of vectors $\{v_0, v_1, \ldots, v_p\}$ where

$$\begin{aligned} v_0 &= x_0 \\ v_1 &= x_1 - \frac{x_1 \cdot v_0}{v_0 \cdot v_0} v_0 \\ &\vdots \\ v_p &= x_p - \frac{x_p \cdot v_0}{v_0 \cdot v_0} v_0 - \frac{x_p \cdot v_1}{v_1 \cdot v_1} v_1 - \ldots - \frac{x_p \cdot v_{p-1}}{v_{p-1} \cdot v_{p-1}} v_{p-1} \end{aligned}$$

Further we define an orthonormal set of vectors $\{u_0, u_1, \ldots, u_p\}$ where $u_i = \frac{v_i}{||v_i||}$ and $||v_i|| = \sqrt{v_{i,1}^2 + v_{i,2}^2 + \ldots + v_{i,J}^2}$ for $i = 0, 1, \ldots, p$.

We will now give a more detailed version of the example given in section 4.2.2. Consider the estimated birth cohorts rates for cohort c as

$$\hat{\gamma}_{c-6},\ \hat{\gamma}_{c-4},\ \hat{\gamma}_{c-2},\ \hat{\gamma}_c,\ \hat{\gamma}_{c+2},\ \hat{\gamma}_{c+4},\ \hat{\gamma}_{c+6}.$$

The estimated rates are given for 2-year age and period intervals and corresponds to (4.3). We consider

$$\gamma_{c+2j} = \theta x_{0j} + \eta_1 x_{1j} + \eta_2 x_{2j} \tag{C.1}$$

for $j = 0,\ \pm1,\ \pm2,\ \pm3$, where $x_{0j} = 1$ is a vector only containing ones, $x_1 = (0, 2, \ldots, 12)'$ and $x_2 = x_1^2 = (0, 4, \ldots, 144)'$. By the definitions given above we find that

$$\begin{aligned} v_0 &= (1\ 1\ 1\ 1\ 1\ 1\ 1)' \\ v_1 &= (-6\ -4\ -2\ 0\ 2\ 4\ 6)' \\ v_2 &= (5\ 0\ -3\ -4\ -3\ 0\ 5)' \end{aligned}$$

Note that by using exactly the definition of v_2 given above we would obtain $v_2 = (20\ 0\ -12\ -16\ -12\ 0\ 20)'$. However it does not matter for the results if we use the former or the latter v_2, as long as they are proportional to each other. Further we notice that $\sum_{j=-3}^{3} v_{1j} = \sum_{j=-3}^{3} v_{2j} = 0$ and $\sum_{j=-3}^{3} v_{1j}v_{2j} = 0$. Thus the vectors v_0, v_1 and v_2 are orthogonal. Now the model can be written as

$$\gamma_{c+2j} = \theta v_{0j} + \eta_1 v_{1j} + \eta_2 v_{2j}.$$

However it is common to standardize the weights and obtain an orthonormal set of vectors with length 1, i.e. $u_0 = v_0\frac{1}{\sqrt{7}}$, $u_1 = v_1\frac{1}{\sqrt{112}}$ and $u_2 = v_2\frac{1}{\sqrt{84}}$. We now replace the orthogonal- with the orthonormal vectors and consider $\gamma_{c+2j} = \theta u_{0j} + \eta_2 u_{2j}$. Thus we can estimate η_2 as in (4.6). When the orthonormal set of vectors are used we will for simplicity define the contrast as K_o. Fortunately the calculations of u_0, u_1 and u_3 are easily obtained in function *poly* in the software R (R Development Core Team, 2010).

Appendix D

R-code

contr.sec()

This function was provided by Bjørn Møller and calculates the estimated second differences. The function can be implemented in the software **R** (R Development Core Team, 2010).

```
contr.sec<-function(n,contrasts=TRUE) {
if(is.numeric(n) && length(n) == 1)
  levs<- 1:n
else {
  levs <- n
  n <- length(n)
}
n <- n-2
#only n-2 second differences
contr<-diag(n)
for(i in (n-1):1) {
contr<-contr+(n-i+1)*cbind(rbind(matrix(rep(0,(n-i)*i),n-i,i),diag(i)),matrix(rep(0,(n-i)
}
contr<-rbind(matrix(rep(0,n*2),2,n),contr)
if(contrasts) {
if(n<2)
stop(paste("Contrasts not defined for", n-2, "degrees of freedom"))
}
contr
}
```

Bibliography

Freddie Ian Bray. *Temporal studies of cancer occurence and applications of the age-period-cohort method to trend in Europe*. PhD thesis, London School of Hygiene and Tropical Medicine, University of London, 2005.

Cancer Registry of Norway. `http://www.kreftregisteret.no/no/Generelt/Om-Kreftregisteret/Historie/`, November 2010a.

Cancer Registry of Norway. `http://www.kreftregisteret.no/no/Registrene/Kreftstatistikk/`, November 2010b.

Cancer Registry of Norway. `http://www.kreftregisteret.no/no/Generelt/Fakta-om-kreft/Kreftsykdommer/Testikkelkreft/`, November 2010c.

D. Clayton and E. Schifflers. Models for temporal variation in cancer rates. I: age-period and age-cohort models. *Statistics in Medicine*, 6:449–467, 1987a.

D. Clayton and E. Schifflers. Models for temporal variation in cancer rates. II: age-period and age-cohort models. *Statistics in Medicine*, 6:469–481, 1987b.

P. de Jong and Gillian Z. Heller. *Generalized Linear Models for Insurance Data*. Cambridge University Press, 2008. ISBN 978-0-521-87914-9.

Carsten Heuer. Modeling of time trends and interactions in vital rates using restricted regression splines. *International Biometric Society*, 53(1):161–177, 1997.

T. R. Holford. Understanding the effects of age, period and cohort on incidence and mortality rates. *Annu. Rev, Pub. Health*, 12:425–457, 1991.

David C. Lay. *Linear Algebra and its applications*. Pearson Education, 2006. ISBN 0-321-28713-4.

R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2010. URL `http://www.R-project.org`. ISBN 3-900051-07-0.

E. Svensson, T. Grotmol, G. Hoff, F. Langmark, J. Norstein, and S. Tretli. Trends in colorectal cancer incidence in Norway by gender and anatomic site: an age-period-cohort analysis. *European Journal of Cancer Prevention*, 11:489–495, 2002.

R. E. Tarone and K. C. Chu. Evaluation of birth cohort patterns in population disease rates. *American Journal of Epidemiology*, pages 85–91, 1996.

R. E. Tarone and K. C. Chu. Nonparametric evaluation of birth cohort trends in disease rates. *Journal of Epidemology and Biostatistics*, pages 177–191, 2000.

S. Tretli and M. Gaard. Lifestyle changes during adolescence and risk of breast cancer: an ecologic study of the effect of World War II in Norway. *Cancer Causes and Control*, 7:507–512, 1996.

E.H. Wanderås, S. Tretli, and S.D. Fosså. Trends in Incidence of Testicular Cancer in Norway 1955-1992. *European Journal of Cancer*, 12:2044–2048, 1995.

B. Yoav and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, pages 289–300, 1995.