Method to estimate cancer overdiagnosis with prostate screening

Method to estimate cancer overdiagnosis with prostate screening

By

Jiarui Hu

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfillment of the Requirements for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2018) McMaster University

(Statistics) Hamilton, Ontario

TITLE: Method to estimate cancer overdiagnosis with prostate screening

AUTHOR: Jiarui Hu

SUPERVISOR: Stephen Walter

NUMBER OF PAGES: 64

Content

List of Figures

List of Tables

Acknowledgement

My greatest gratitude goes to my supervisor Dr. Stephen Walter who has helped me a lot with his expertise, encouragement and guidance during my journey in this thesis.

I would like to thank my thesis committee members: Dr. Shui Feng and Dr. Roman Viveros-Aguilera for their comments, questions and feedback.

I would like to thank the Department of Mathematics and Statistics at McMaster University for providing a gracious learning environment.

Most importantly, I want to thank my family, especially my parents, for their love and support that have helped me to continue my studies.

Abstract

Aim: Several studies have tried to quantify overdiagnosis of prostate cancer with Prostate-specific antigen(PSA) screening, but estimates vary widely. This study aims to evaluate the degree of overdiagnosis of prostate cancer with 10 or 14 follow-up years after the stop of screening in Finland.

Methods: We selected 80379 men aged 55-69 years who were randomized to a screening or a control arm, distinguishing four birth cohorts: 1941-44,1937-40, 1933-36 and 1929-32. The first PSA screening test occurred during1996-1999. Men without detected as prostate cancer in the previous screening would be invited to the next screening 4 years later. The estimate of overdiagnosis is the ratio of the cumulative excess incidence to the cumulative incidence of prostate cancer in the screened group after the year-specific incidence became stable.

Results: The patterns of all incidences in these four cohorts have not become stable yet, and the difference of cumulative incidence in the current longest follow- up years is the best estimate of overdiagnosis so far.

Conclusion: Overdiagnosis rates of prostate cancer in people who received screening in Finland was estimated as 2.27%,15.4%, 11.4%, and 10.2% for 1929-32, 1933-36,1937-40, and 1941-44 cohorts, respectively.

Chapter 1

Introduction

Prostate cancer(PCa) is the third-leading cause of death from cancer in males. Different regions have varying incidence and mortality. The risk of PCa is 74% higher in blacks than in whites but remains low in Asians. Considering the experimental conditions, clinical and biopsy studies mostly focus on the people in western countries.

In the US, the most commonly diagnosed cancer in men is prostate cancer. The American Cancer Society (ACS) estimated that during 2017, about 161,360 new cases of PCa would be diagnosed in the US with an estimated 17% death rate ("Cancer Facts & Figures 2017", 2018).

Regarding Canada, it was estimated that 21,300 men would be diagnosed with prostates cancer, which represented 21% of all new cancer cases in males, and 4100 men which represented 10% of all cancer deaths in men would die from prostate cancer in 2017.

The survival time of PCa has a close relationship with the extent of tumor at the time of diagnosis. Therefore, a screening program that could sensitively detect aggressive localized tumors in asymptomatic men might lead to a dramatic decrease in mortality of PCa. PSA testing revolutionized this screening.

In 1986, PSA testing was first approved by Food and Drug Administration(FDA) to monitor the progression of PCa. In 1994, FDA approved the use of PSA to test PCa for asymptomatic men. So, incidence rates for PCa had a dramatic increase in the late 1980s and early 1990s led by widespread screening with the PSA test. The infrequency of PCa for people below 40 implied that men should make an informed decision about PSA test (Smith RA, 2018). Moreover, ACS (American Cancer Society) guideline updated in 2001 indicated there was still uncertainty about the overall value of periodic testing when associated with a reduced risk of death from PCa. Comparison of reduction in PCa mortality in the US and Europe found that a study conducted in the United States failed to find any mortality benefit compared with the European study, where a 21% reduction was demonstrated. (Schröder et al., 2009)

As a result, ACS began to recommend that it was not necessary to take PSA testing for asymptomatic men who had less than a 10-year life expectancy, and physicians were required to provide detailed information about risk and potential harms of early detection.

Although the benefits of PSA testing remain controversial, people had a long concern on the adverse effect of PSA testing mainly because of overdiagnosis, which was the detection through screening of cancer that would have never been identified in the absence of screening (Etzioni et al., 2018). Overdiagnosis in cancer screening could result from slow growth of the tumor or mortality before cancer would have caused symptoms, in which

cases, there would only be harmful effects on people if screening found cancer that would have never become clinically detected. Because it is impossible to recognize which individual case of cancer is the results of overdiagnosis, investigators need to find methods to quantify overdiagnosis associated with cancer screening indirectly, but there is a significant variation in these estimates.

In general, there are two main approaches to estimate the overdiagnosis rate: modeling of disease transition and the excess-incidence approach. (Etzioni, R., & Gulati, R., 2015)

The first approach models the pattern in which PCa will hypothetically occur without screening, and the trend by which cancer occurs with screening, then comparing these two models to calculate the rate of overdiagnosis, such as MISCAN (Draisma & de Koning, 2003; Draisma et al., 2009), which is a microsimulation model that simulates individual life history as a Markov process of states and transition to calculate overdetected rate by deriving the lead time, UMich model (Tsodikov, Szabo & Wegelin, 2006),where a statistical model was used to capture the features of PCa incidence registered for prostate cancer then helped us to predict lead time and overdiagnosis of prostate cancer, and FHCRC model (Etzioni et al., 2007; Gulati, Inoue, Gore, Katcher & Etzioni, 2014) where a microsimulation model linked individual's PSA levels with the progression of prostate cancer.

In all these simulating models above, investigators need to find the balance between the complexity and the transparency of the model. The complexity of

the model can be adjusted from simple, involving only a few features of the disease to complex, referring to many features and many transitional probabilities. Therefore, the disadvantage of simulating method is apparent. If the complex model is used to capture as many features of the disease as possible, it is difficult to evaluate potential biases of the result due to lacking in transparency. Otherwise, the simplicity may be not able to reflect the natures of the original data completely.

The second approach uses the observed excess incidence rates—the difference between the screening and the control group. We regard this as the “excess-incidence” approach. The difficulty of this approach is to observe the counterfactual incidence data in the absence of screening.

In this study, the second approach was chosen to calculate the overdiagnosis rate, and counterfactual incidence data was taken from a randomized clinical trial. In this trial, men were randomized into two groups, the one in the screened group were offered PSA testing, men in the other group would not be screened during the same period. Ideally, it was assumed that there were similar underlying risks of PCa in the two groups.

The estimation of overdiagnosis was assumed to be the difference of cumulative differences between the screening and the control group, because research suggested that calculating the estimation from cumulative incidence with data taken from randomized trials was the most valid method when there was a sufficient follow-up after the last screening (Biesheuvel, Barratt, Howard, Houssami & Irwig, 2007). In particular, excess incidence in a stop-

screen trial required using cumulative incidence, whereas a continued-screen trial or population setting required using annual incidence. (Gulati, Feuer & Etzioni, 2016) Therefore, we used the cumulative incidence to calculate the frequency of overdiagnosis for the stop-screen trial of Finland.

Chapter 2

Data and Method

2.1 Data

In this analysis, all data was taken from Finland section of the European Randomized study of Screening for Prostate Cancer (ERSPC). The ERSPC is a multi-center, randomized screening trial between an intervention arm offered PSA screening and a control arm without any intervention.

The Finland section of the ERSPC, one of eight participating countries of the ERSPC, began in 1996. About 80379 men aged 55-69 years were randomized to a screening or a control arm, distinguishing four birth cohorts: 1941-44,1937-40, 1933-36 and 1929-32. The men in the screening group were offered to screen every four years from their first screening time. The first round was performed in 1996-99, the second round in 2000-2003. The final round occurred in 2004-2007 but excluded men aged >71 years. A PSA level 4.0 ng/ml was used as the indication for biopsy. Also, the digital rectal examination was offered to the men with serum PSA between 3.0 ng/ml and 3.99 ng/ml.

2.2 Definition of “catch-up” point

For a randomized screening study, Figures 2.1 and 2.2 display the trend of incidence in the presence or absence of overdiagnosis, respectively.

Each curve (Figures 2.1 or 2.2) of year-specific incidence in screening group showed a peak during every screening round. These increases could be explained by the fact that cancer which would have been presented clinically in later time were diagnosed by PSA screening. After the screening stopped, the year-specific incidence of screening group was lower than the incidence of control group because cancer that would have been diagnosed in these years had already been detected earlier by screening. If there was overdetection (Figure 2.1) due to the screening, there would be an excess of incidence of PCa in the screened group compared with the unscreened group. Otherwise (Figure 2.2), the cumulative incidence in the screened and control group would equalize once the lead time was accounted for.

The frequency of overdetection is all the excess incidence once lead-time is accounted for. Lead time is the time interval from screen detection to the time of clinical presence. When the lead time has elapsed, the year-specific incidence in the screening group is expected to catch up with the incidence in the unscreened control group.

We define the point at which year-specific incidence of PCa in the screening group equalizes the rate in the control group as the “catch-up” point. Then, the difference of year-specific rate is used to build a spline regression model to determine the value of “catch-up” point. Compared to the cumulative one, the

independence of year-specific incidence difference offered a prerequisite to establishing regression model. Moreover, it was much easier to use the spline regression model for defining whether the incidence became stable (slope=0) than any time series model, although spline regression model's loss of smooth might lead to the poor performance of fit to data. To give an insightful evaluation of the spline regression model and accurate estimation of the "catch-up" point, we simulated the original data before we fitted the regression model.

Figure 2.1: Schematic plot for the men offered PSA screening at 60, 66, and 71 years old (based on the hypothetical data) in the presence of overdiagnosis

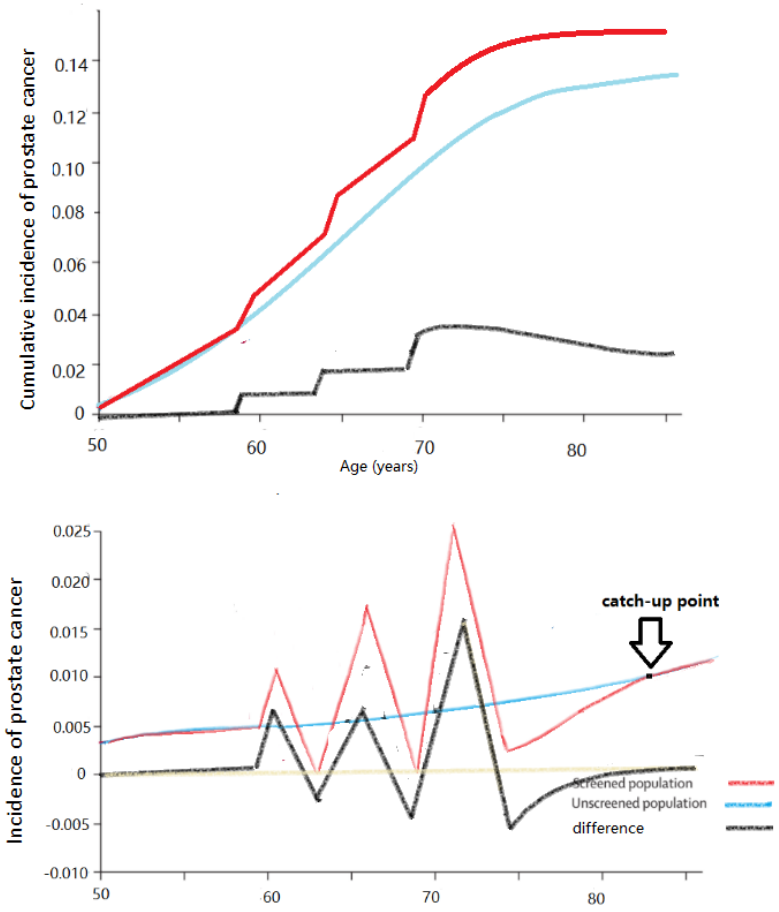
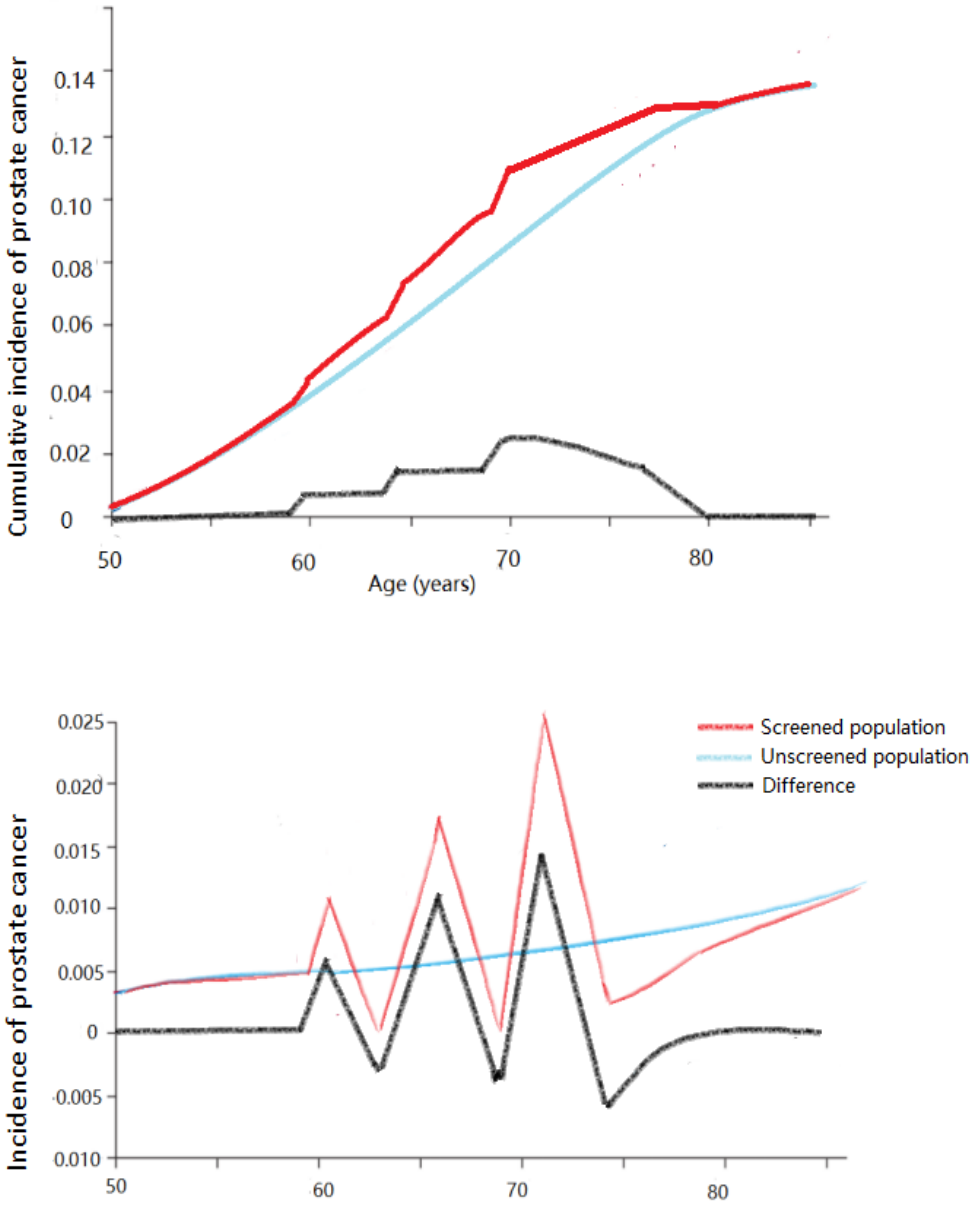


Figure 2.2: Schematic plot for the men offered PSA screening at 60, 66, and 71 years old (based on the hypothetical data) in the absence of overdiagnosis



2.3 Simulation for finding "catch-up" point

Simulation is a technique for predicting the performance of experiment on the model of the system. The experiments are calculated using a computer model rather than on the real system as the latter will be ineffective, expensive and time-consuming. In this thesis, a simulation study was conducted to evaluate the availability of finding the stabilized point by using the spline regression model.

We obtained a random sample of observed data from the density of the normal distribution with pre-specified values of parameters N(μ, σ). The year-specific rate difference was regarded as the parameter μ, while σ could be calculated as follows:

The year-specific rate of different age cohort was assumed to follow the Poisson distribution. So, the estimated sample variance of the year-specific rate could be calculated by:

$$Var(\mu_{ij}) = \frac{x_{ij}}{{\eta_{ij}}^2} \qquad i = 1,2 \qquad (1)$$

Where $x_{1.}$ is the number of PCa cases in the screening group, η_{1j} is the total number of people in the screening population at the beginning of the j-th visit, μ_{1j} indicates the j-th year-specific incidence of screening group. x_{2j} ,η_{2j} and μ_{2j} are corresponding variables in the control group.

Since the control group and the screening group were independent, the variance of the rate difference could be calculated by

$$Var(\lambda_j) = Var(\mu_{1j}) + Var(\mu_{2j}) \quad (2)$$

Where λ_j represents the j-th year-specific incidence difference.

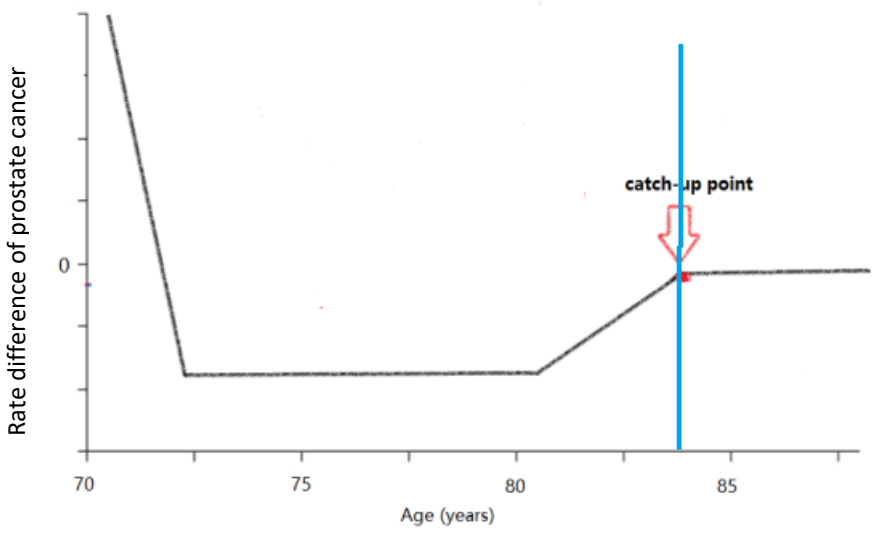
According to the central limit theorem, the year-specific rate difference followed a normal distribution.

Since the “catch-up” point could only occur after the last screening, we simulated year-specific rate difference data after last screening to follow a normal distribution N (μ, σ) for 100 or 500 times with μ equal to the rate difference and σ calculated by equations (1) and (2).

To find the “catch-up” point, spline regression was applied to simulate year-specific rate difference of different age cohort between screening and control group after the last screening.

Ideally, the spline regression model for the simulated data will approximately have the trends presented in Figure 2.3. The sharp decrease at the beginning is because cancers that would have been diagnosed for the screening group in these years had already been detected earlier by screening and so incidence of the screening group appears lower than the control group for a few years. Eventually, at some point in time after screening stops, the incidence of PCa in the screened and unscreened groups will be equal. Catch-up point will be the point after which there is no longer any difference in annual incidence between the control group and screening group.

Figure 2.3: Schematic plot of spline regression model with four segments for year-specific rate difference.

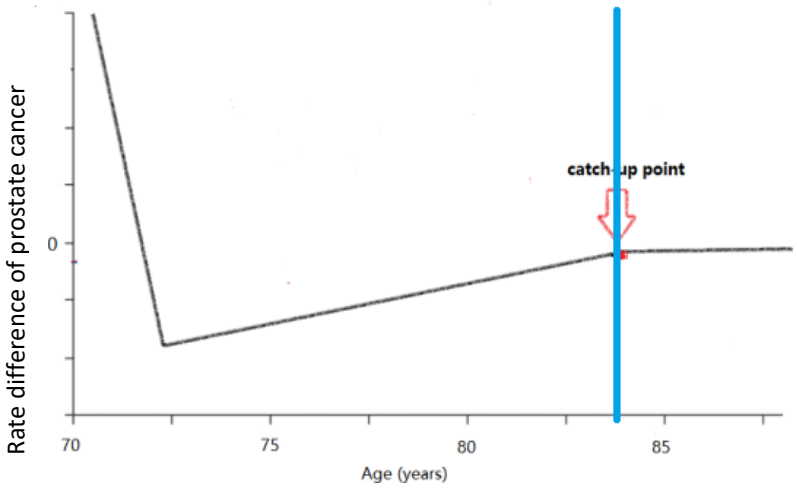


The part of the graph before the blue line displays the model when follow-up years are not long enough for rate difference to become stable.

This model can be achieved if there is enough data point for every segment, especially for the second and third segment of this model.

Otherwise, Figure 2.4 shows a compromise model without enough data points for the second and the third segment. “Catch-up” point has the same definition as the model with four segments.

Figure 2.4: Schematic plot of spline regression model with three segments for the simulated year-specific rate difference.



The graph before the blue line displays the model when follow-up years are not long enough for rate difference become stable.

In this study, Akaike Information Criterion (AIC) was chosen to select the number of breakpoints and the best fitted spline regression model among 100 models created by simulating. AIC could estimate the relative loss-information by dealing with the trade-off between the goodness of fit of spline regression model and the complexity of the model. The established model could then give an estimate of the value of breakpoints and slopes in each segment.
For every breakpoint, an initial value was required to input to build the spline regression model. The initial value of each breakpoint was determined by the trend of year-specific rate difference of PCa. Figure 2.3 and 2.4 explicitly displayed that the first break point occurred when year specific rate reached the lowest point. Regarding Figure 2.3, the second breakpoint is exactly the "catch-up" point when the rate difference started becoming stable. For a model like Figure 2.4 with four segments, the second breakpoint is the point when the rate difference began to increase, and the third one is the "catch-up"

point. Therefore, the first original breakpoints value was set at 9th and 17th follow-up years for the 1941-44,1937-40, and 1933-36 cohort, 7th and 12th follow-up years for the 1929-32 cohort.

The prerequisites for the complete model with “catch-up” point (as Figures 2.3 and 2.4) is enough follow-up years as long as the longest lead time for screening being fully adopted. (Gulati, Feuer & Etzioni, 2016).
However, the estimation of duration of the detectable preclinical phase (DPCP), which is the upper limit of lead time, is 10-14 years (Auvinen A, 2018) longer than our current longest follow-up years for the Finland data. So, the problem about how to verify the availability of finding catch-up point by using spline regression need to be resolved when follow-up years of original data are not long enough. In such condition, artificial data was applied to prove whether spline regression could determine the catch-up point when the follow-up years of real data are not long enough.
First, we set up a segmented inputting function (as Figures 2.3 and 2.4) based on the model fitted by the original Finland data but allowing for a long period of follow-up years for the year-specific rate difference to come back to zero and to become stable. Secondly, we simulated inputting function to follow a normal distribution, the sample variance of which was taken from the variance of real data. Finally, spline regression was used to fit these artificial data to verify whether we could estimate the value of “catch-up” point when its existence was surely confirmed.

2.4 Estimation of Overdiagnosis

The definition of overdiagnosis used in this analysis was screening-detected cancer that wouldn't have been clinically significant for the remainder of the patient's life in the absence of screening. Therefore, the frequency of overdetection was all the excess incidence once "catch-up" point was confirmed. Then the measure of overdiagnosis rate could be calculated as follows:

$$\left(\frac{\begin{matrix}Cumulative\ incidence\ of\ prostate \\ canceer\ in\ screened\ population\end{matrix} - \begin{matrix}Cumulative\ incidence\ of\ prostate \\ cancer\ in\ unscreened\ group\end{matrix}}{\begin{matrix}Cumulative\ incidence\ of\ prostate\ cancer \\ in\ screened\ group\end{matrix}}\right) \times 100\% \quad (3)$$

The 95% confidence interval for the frequency of overdiagnosis could be calculated as follows:

$$I_s - I_c \pm 1.96\sqrt{s_s^2 + s_c^2} \quad (4)$$

Where I_s, I_c, and s_s, s_c are cumulative incidence and standard error of the screened population and unscreened population, respectively.

Chapter 3

Results

3.1 Prostate cancer Incidence

Data used in this study was taken from the Finland data, consisting of 80,458 men born from 1929 to1944, which was part of European Randomized Study of Screening for Prostate Cancer (ERSPC).

Figures 3.1,3.2,3.3, and 3.4 displayed the prostate cancer incidence for 1929-32, 1933-36, 1937-1940, and 1941-1944 cohorts.

A preliminary glance at the plot of prostate cancer incidence showed there were apparent differences among the different birth cohorts. The final cumulative incidence value in Figures 3.1, 3.2, 3.3, and 3.4 demonstrated the fact that the prevalence had a close relationship with age. (Bell, Del Mar, Wright, Dickinson & Glasziou, 2015). Based on above fact, all the analysis and results were decided to be displayed by different birth cohorts.

The trend of all these plots verified our assumption set in Figure 2.1. Specifically, the difference of cumulative incidence for the 1929-32 cohort seemed to trend closely towards zero-difference between the screening group and the control group, whereas the other cohorts still retained a non-zero difference.

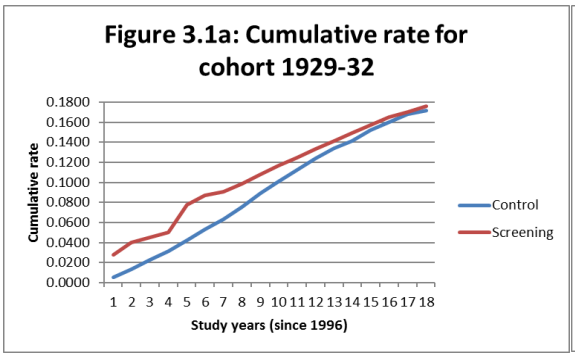
For Figure 3.1, the two peaks in incidence of screening corresponded to the two screening rounds required by the study protocol. The first screening occurred during 1996-1999, the second was 2000-2003 (corresponding to the X-axis 1 and 5). When participants in the screening group stopped undergoing screening at year 5, its year-specific incidence fell below control-arm

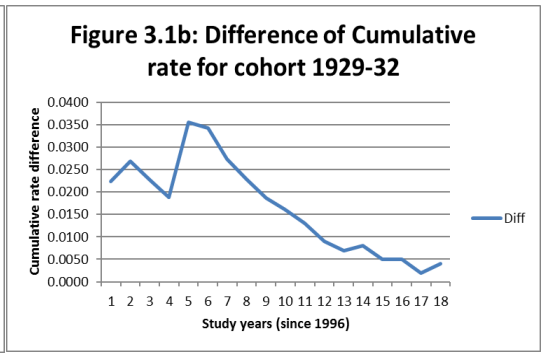
incidence because of lead-time effect. The year-specific incidence in the screening group then gradually rose back to incidence of control group. The annual incidences of two groups seemed to tend closely to zero-difference, which implied that excess cumulative incidence first might be equal to the number of overdiagnosed cancer at about 18 years.

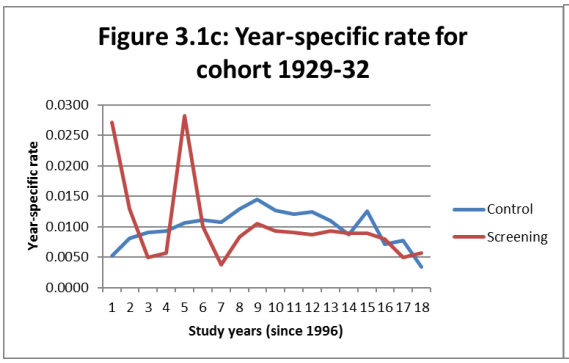
Three humps in the incidence of screening group (Figures 3.2, 3.3, and 3.4) corresponded to the three screening rounds required by the study protocol. The first screening occurred in 1996-99, the second was 2000-03, and the third was 2004-07. (corresponding to the X-axis 1, 5 and 9). After screening stopped at 9 years, the year-specific incidence of screening group was less than the incidence of the control group for several years because of lead-time effect. The year-specific incidence of screening group then began to come back to control-arm incidence gradually.

In Figures 3.1B, 3.2B, 3.3B, and 3.4B the cumulative excess rates of prostate cancer were shown for different age during the study years. The excess incidence got the highest point during the last screening round. At the end of the follow-up, the cumulative excess rate was about 4.0e-03 for men enrolled in the program when they were 67-70 years old. For men aged 55-58 and 59-62 years at the enrolment, the excess cumulative incidence decreased over time to 1.5e-02 and 1.0e-02 after the last screening, respectively. Unlike other cohorts, the excess incidence of men aged 63-66 years at the enrolment fluctuated and finally dropped to 2.5e-02 at the end of follow-up years.

Figure 3.1: Prostate cancer incidence of cohort 1929-32.







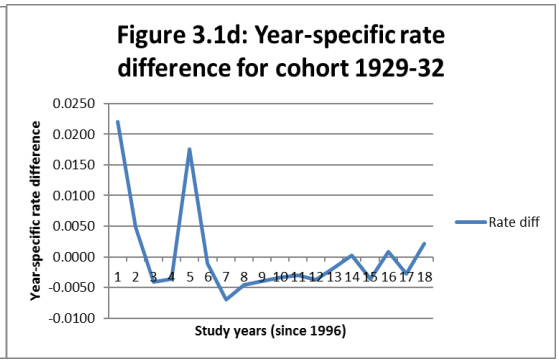
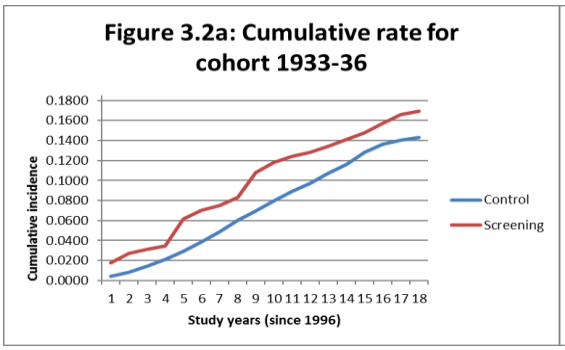
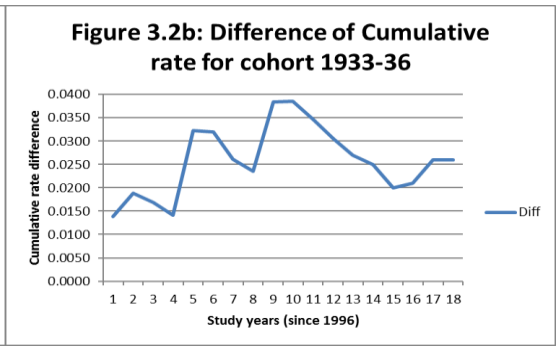
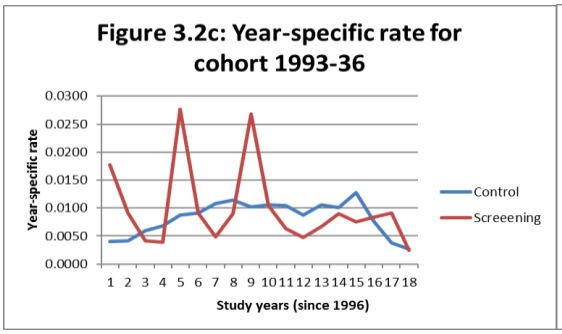


Figure 3.2: Prostate cancer incidence of 1933-36 cohort.







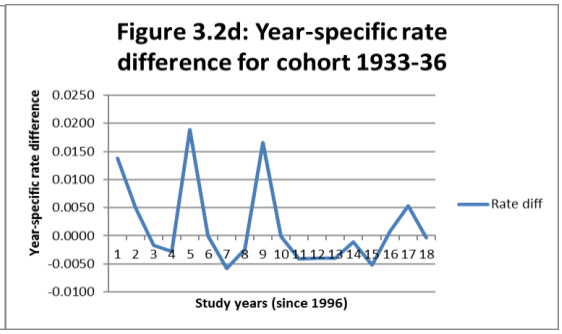
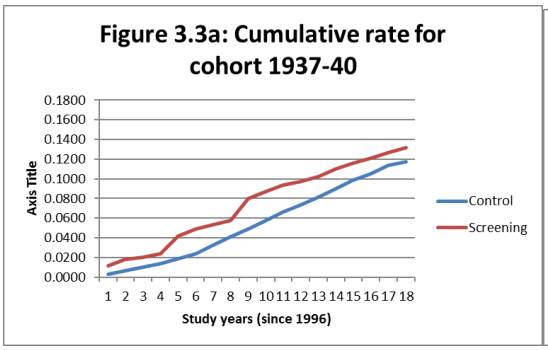
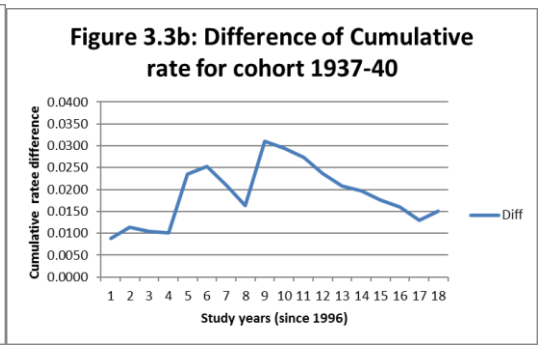
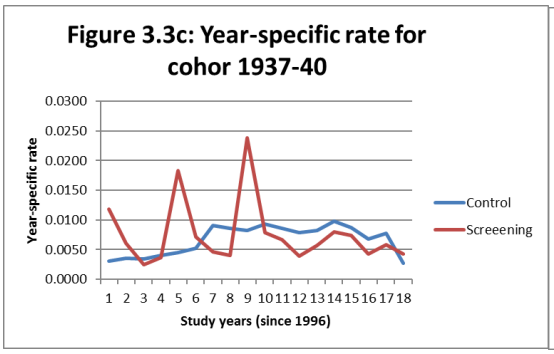


Figure 3.3: Prostate cancer incidence of 1937-40 cohort.







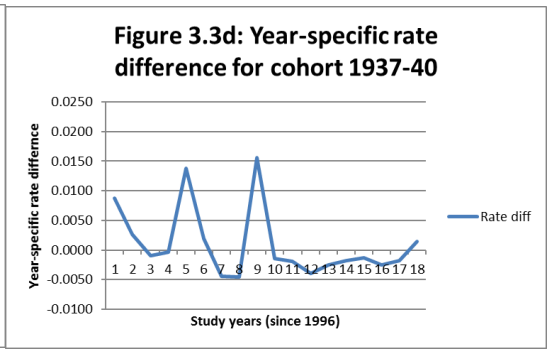
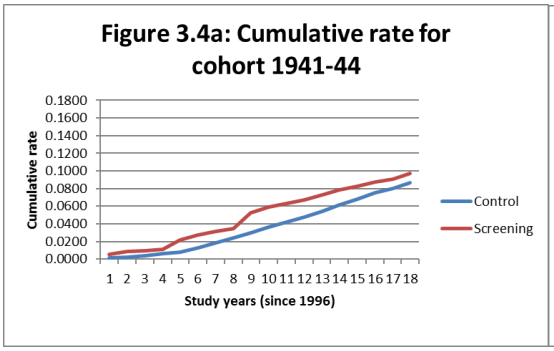
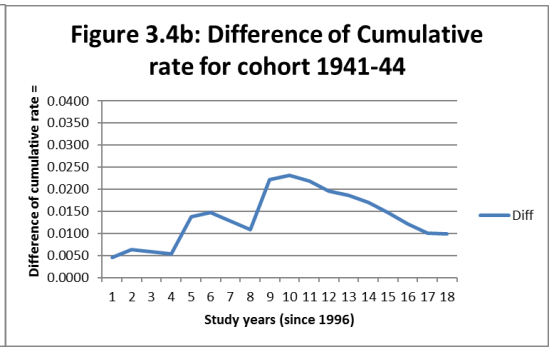
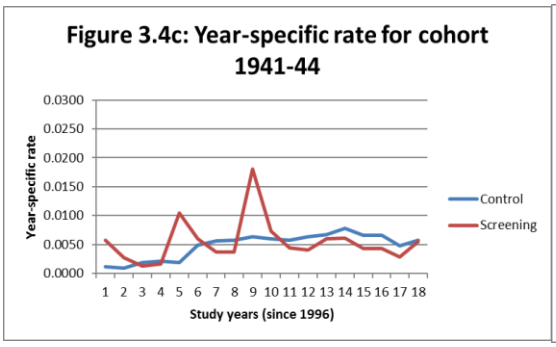
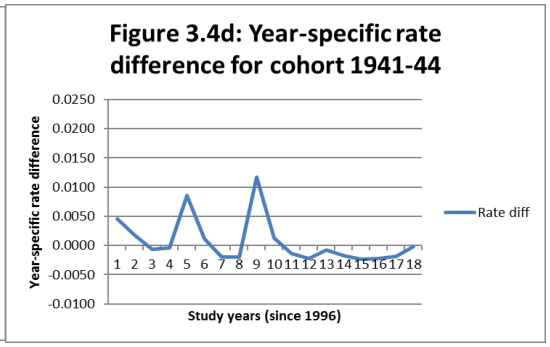


Figure 3.4: Prostate cancer incidence of 1941-44 cohort.









Figures all showed an obvious increase in incidence in the screened group during their first screening round. This increase still happened in the following screening rounds, because it was difficult to get the perfect sensitivity level of screening and there were still new cases that had developed since previous screening. After screening stopped, the year-specific incidence of screening group continued decreasing until reaching the lowest point. Then the year-specific rate difference maintained the same level for several years and then reverted to 0.

According to the trend of the year-specific rate difference, the lowest point and the point at which rate difference began to increase and after which rate difference was equal to 0 were decided to be the initial value of breakpoints to fit the spline regression model iteratively. So, the 11, 15 and 17 follow-up years were set to be the original joint points for 1933-36,1937-40 and 1941-44 cohort. For 1929-33 cohort, the original breakpoints were chosen to be 7, 12 and 14 follow-up years.

3.2 Goodness of fit of Spline regression model

AIC was chosen to be the criterion to assess the fit of the predicted model with various numbers of breakpoints.

Table 3.1: Summary of the AIC value of spline regression model with different number of breakpoints.

AIC	1929-32	1933-36	1937-40	1941-44
1 break point	-135.2	-84.6	-100.4	-94.4
2 break point	-132.7	-81.1	-104.8	-104.8
3 break point	-129.1	NA	NA	NA

NA: Not converged when fitting the initial data to spline regression model.

The smallest AIC value suggested that this model had the best fit and lost minimum information compared to the model with larger AIC value. According to Table 3.1, appropriate number of breakpoints for the different cohort was chosen for the next simulating process: 1 joint point for 1929-32 cohort and 1933-36, 2 for 1937-40 and 1941-44 cohort.

Given the proper number of breakpoints we need to estimate for different age cohort, the model also required an initial inputted value for every breakpoint to fit model.

For cohort 1929-32: First, we calculated the sample variance of rate difference, then simulated the real rate difference 100 times by assuming they followed a normal distribution. Secondly, we fitted the simulated data to a spline regression model. Table 3.1 implied that the model with only one break point was the best spline regression model. Figure 2.3 displayed that the breakpoint of the model with two segments was exactly the point when rate

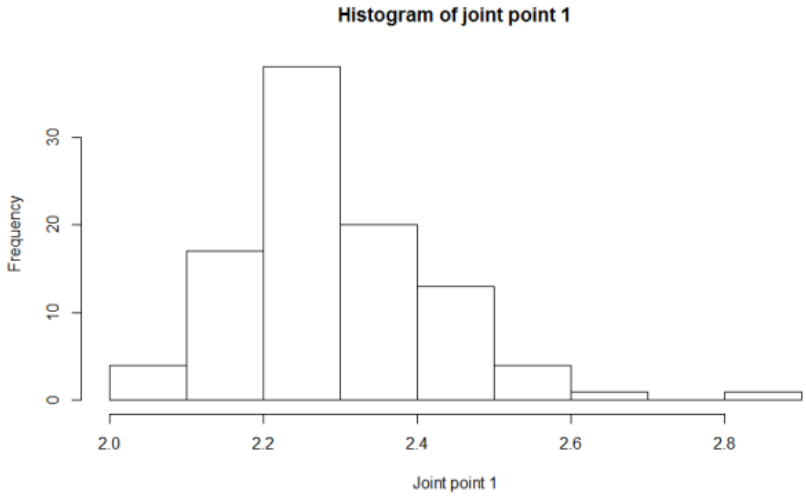
difference attained its lowest value. We preliminary assumed this happened at 7 follow-up years, so we set 3 (corresponding to 7 follow-up years, since we only simulated data after 5 follow-up years, the last screening time point) as the initial value of break points to fit spline regression model. Finally, we estimated the value of breakpoint and slope3 by analyzing statistics after fitting the simulated data.

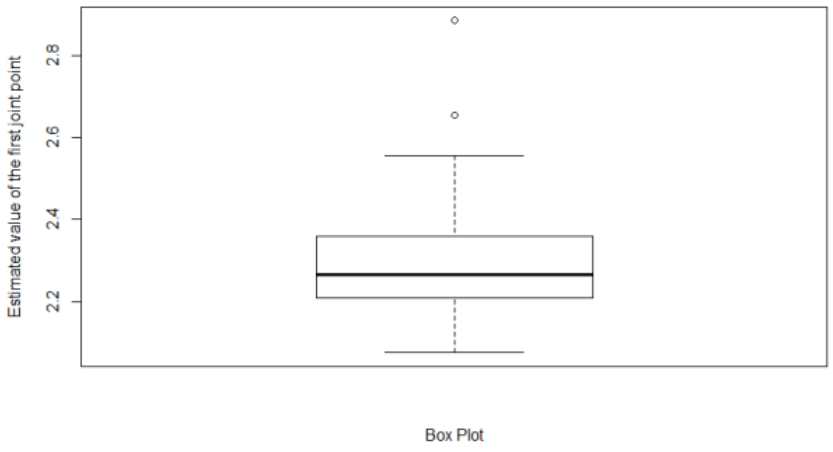
Among the 100 times simulated data, 98% percentage of simulated data succeeded in fitting to the spline regression model with 1 joint point. All statistics of parameters were reported in Table 3.2. Mean value was selected to estimate the value of breakpoint and slope for the second segment because the histogram plot denoted that they roughly followed a normal distribution when excluding two outliers in the boxplot of joint point 1 (Figure 3.5). Therefore, the estimated value of breakpoint (the point when year-specific rate difference reached to the minimum value) in the model was 2.29 years. In terms of slope2, the estimated value was 5.9e-4. The histogram plot (Figure 3.5) of slope2 illustrated that 0 is not contained in the whisker range of slope2. This phenomenon told us that the slope of the second segment was significantly larger than 0.

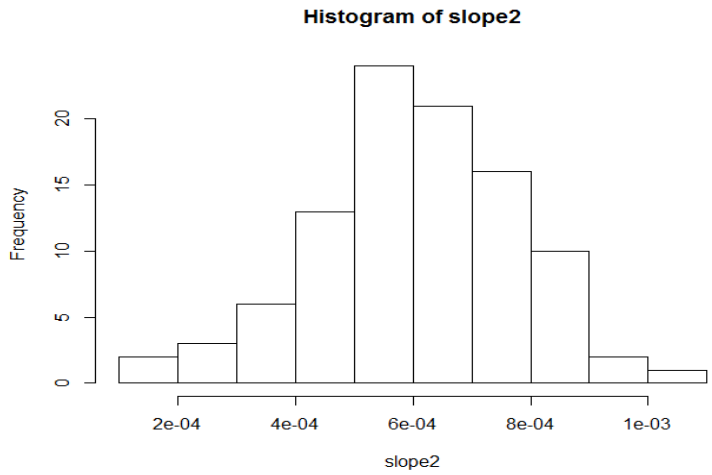
To verify the goodness of fit of this model to real data, we made a plot (Figure 3.6) of models with the smallest AIC (-140.2) value among 100 spline regression models for simulated data. The line chart reached trough at around the second follow-up year since last screening stopped and then began to rise

to 0. The estimated value and whisker range (Figure 3.5) of slope for the second segment implied that the last segment had not become stable yet.

Figure 3.5: The histogram plot and the box plot of the first break point and the slope for the second segment for fitted spline regression model after 100 times simulation for 1929-32 cohort. Box plots include median and interquartile range; there are two points beyond the whiskers range of joint point 1 and three outliers in the distribution of parameter slope2.







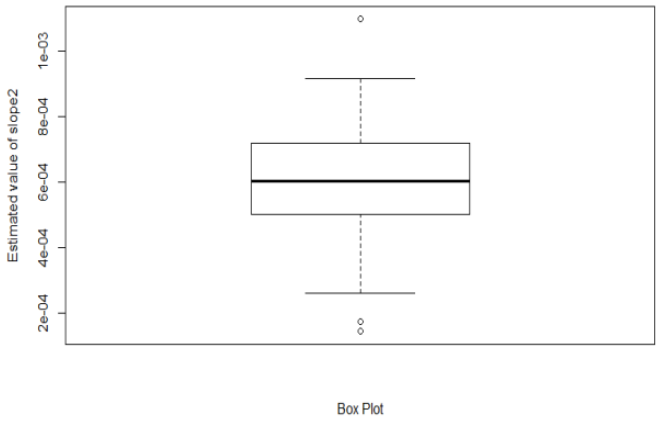


Table 3.2: Summary of statistics data for parameters of spline regression by cohort 1929-32.

para	Joint point 1	Slope1	Slope2
mean	2.29	-0.018	0.0006
sd	0.140	0.00302	0.00019
NA	2%		

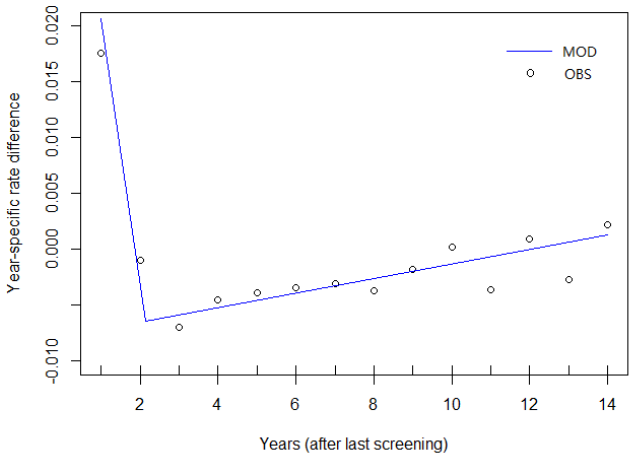
NA: Percentage of cases that are not converged among 100 times simulation.

SD: Standard deviation for every parameter we estimated

Joint point 1: the value of joint points

Slope 1 and 2: slopes for the corresponding segments.

Figure 3.6: Year-specific prostate cancer rate difference for 1929-32 cohort in the Finland section of the ERSPC trial as original observed data (OBS), and as predicted by the spline regression model (MOD). The model has the fewest loss information chosen by AIC value.



For cohort 1933-36: We repeated the same procedure to obtain simulated data for fitting spline regression model as cohort 1929-32. Similarly, the spline regression model had the smallest AIC value. By Figure 3D, a rate difference of prostate cancer fell to the lowest point at 11 follow-up years, so we also set 3 (corresponding to 11 follow-up years) as the first original breakpoint to fit regression model. There was 99% percentage of simulated data that converged to the spline regression model. All statistics of parameters were reported in table 3.3. Mean value was selected to estimate the joint point and slopes because the distributions of breakpoint and slope2 were roughly symmetric and unimodal when not considering outliers. Therefore, we

estimated rate difference of prostate cancer for people born from 1933-36 attained the minimum value at around 2.41 follow-up years after the last screening visit and then began to increase at the slope of 1.01e-3

Figure 3.8 plotted the best spline regression model with the smallest AIC (-101.4) value. In this plot, the lowest point occurred at around 2.5 follow-up years, which corresponded to the estimation above. After that, the rate difference appeared to become stable. However, the box plot showing that 0 was not contained in the whisker range of slope2 indicated the estimated value of slope was significantly larger than 0.

Table 3.3: Summary of statistics for parameters of spline regression by cohort 1933-36.

para	Joint point 1	Slope1	Slope2
mean	2.41	-1.61e-02	1.01e-03
sd	0.240	3.366e-03	2.506e-04
NA	1%		

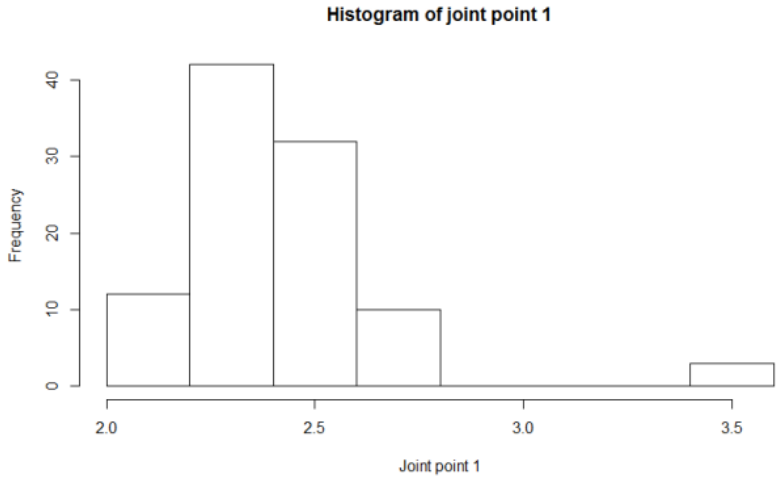
NA: Percentage of cases that are not converged among 100 times simulation.

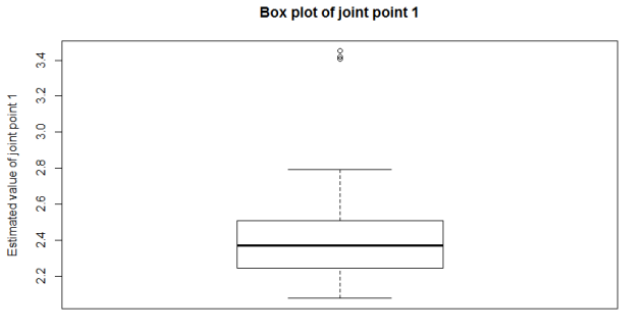
SD: Standard deviation for every parameter we estimated

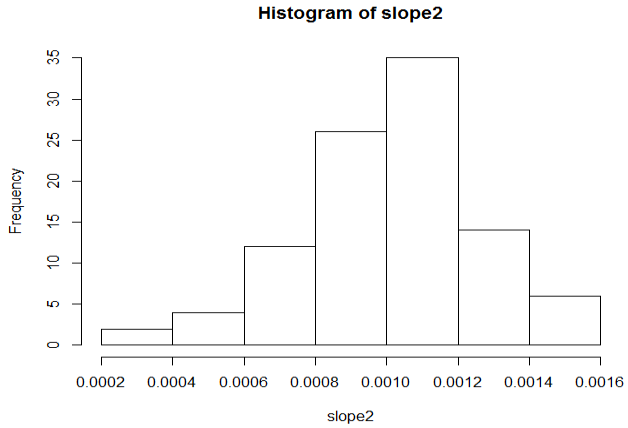
Joint point 1: the value of joint points

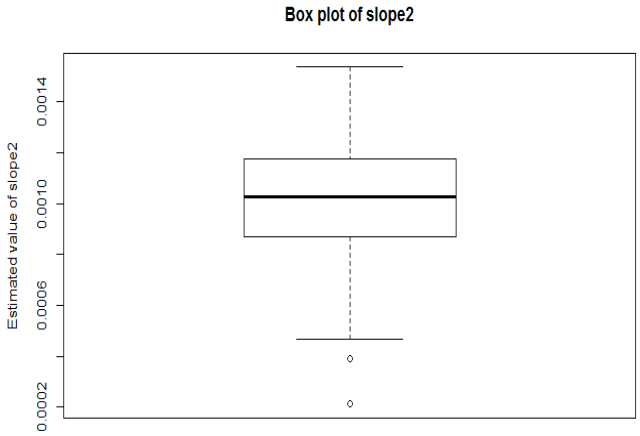
Slope 1 and 2: slopes for the corresponding segments.

Figure 3.7: The histogram plot and box plot of the first break point the slope for the second segment for fitted spline regression model after 100 times simulation for 1933-36 cohort.



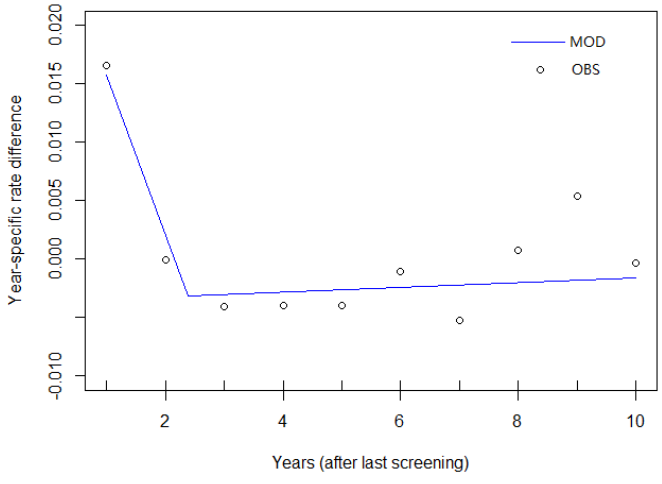






Box plots include median and interquartile range; there are three points beyond the whiskers range of joint point 1 and two suspected outliers in the distribution of parameter slope2.

Figure 3.8: Year-specific prostate cancer rate difference for cohort 1933-36 in the Finland of the ERSPC trial as original observed data (OBS), and as predicted by the spline regression model (MOD). The model has the fewest loss information chosen by AIC value.



Unlike 1929-32 and 1933-36 cohorts, spline regression model with two break points was the best choice for the 1937-40 cohort according to table 3.1. The plot of incidence of prostate cancer displayed that year-specific rate difference fell to the lowest point at the 11th follow-up year and maintained the same level until the 17th follow-up year, at which rate difference began to increase, so we set 3 (corresponding to 11 follow-up years) and 8 (corresponding to 17 follow-up years) as initial value of breakpoints to fit spline regression model. First, we simulated real data 100 times. Among the 100 spline regression models, there was about 40% of simulated data that converged to the spline regression model with three segments. For the other 60% instances, there was only one data point in some segments. Therefore, to acquire enough number of data that converged to this spline regression model, we expanded the sample size for simulation from 100 to 200. All statistics results of parameters were reported in Table 3.4. The histogram plot (Figure 3.9) of breakpoint 1 and breakpoint 2 displayed that distributions of these two parameters were asymmetric. This fact implied that mean value would overestimate or underestimate the value of these two break points. So, the median would be a more appropriate statistic to estimate these two breakpoints. We estimated that year-specific rate difference reached the bottom at around the 2nd follow-up year. Afterward, rate difference started reverting to 0. Since the histogram plot of slope3 was roughly symmetric, it was reasonable to estimate this parameter by using mean value.

Figure 3.10 plotted the best model with the minimum AIC value (-104.70). There were three segments of year-specific rate difference in this model as

displayed in Figure 3.10. Rate difference dropped sharply to the lowest point in the 2nd follow-up year (since last screening time), then almost leveled-off for 7 years. Afterward, rate difference continued increasing until the end of the trial. Obviously, year-specific rate difference had not become stable yet. Under this condition without enough follow-up years, “catch-up” point was unable to be confirmed.

Table 3.4: Summary of statistics for parameters of spline regression by cohort 1937-40.

Para	Joint point 1	Joint point 2	Slope1	Slope2	Slope3
mean	2.14	7.78	-0.016	1.02e-04	2.41e-03
Median	2.10	8.42	-0.016	7.59e-05	2.51e-03
SD	0.119	1.413	0.0025	5.49e-04	1.642e-03
NA	60%				

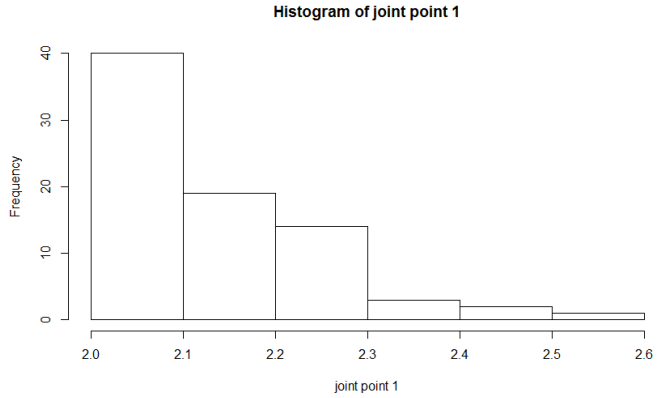
NA: Percentage of cases that are not converged among 200 times simulation.

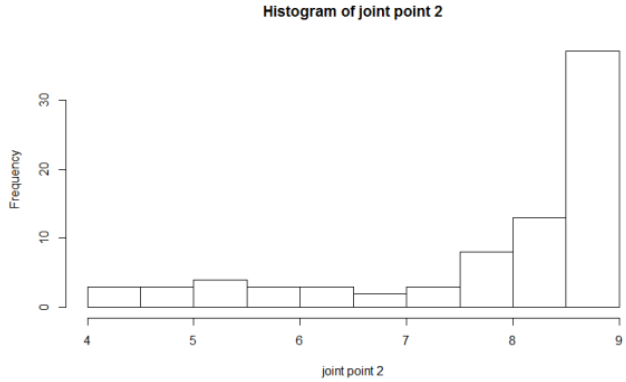
SD: Standard deviation for every parameter we estimated

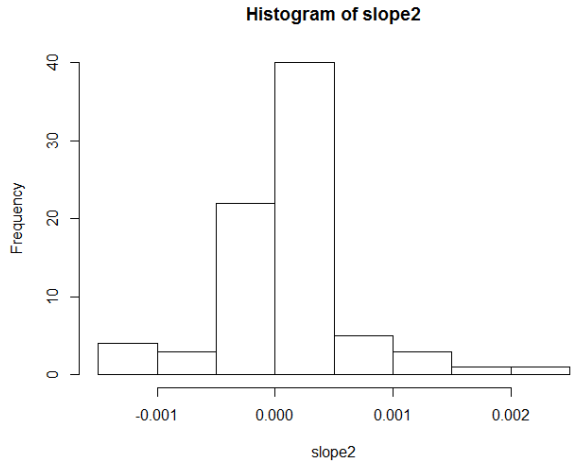
Joint point 1,2: the value of joint points

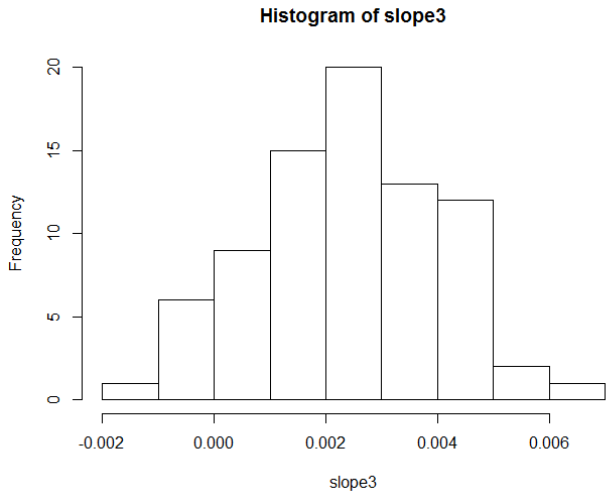
Slope 1, 2 and 3: slopes for the corresponding segments.

Figure 3.9: The histogram plot of breakpoints and slopes for second and third segments of fitted spline regression model after 200 times simulation by cohort 1937-40.



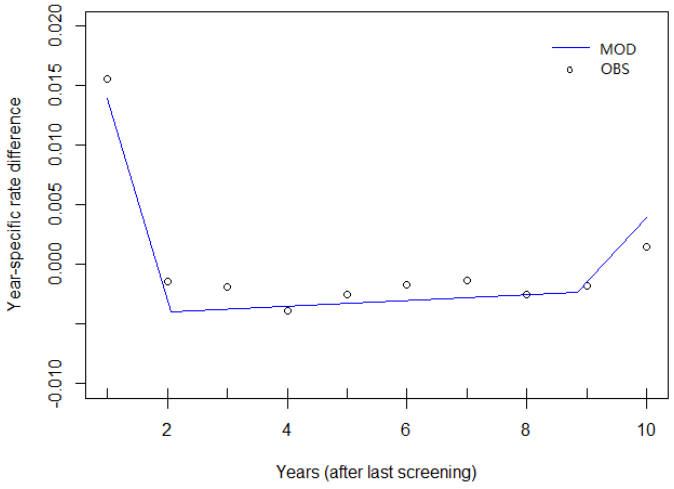






Histogram of Joint point 1 is right skewed, while Joint point 2 is left-skewed. Distributions of parameters are roughly symmetric.

Figure 3.10: Year-specific prostate cancer rate difference for 1937-40 cohort in the Finland of the ERSPC trial as original observed data (OBS), and as predicted by the spline regression model (MOD). The model has the fewest loss information chosen by AIC value



Similarly, spline regression model with two break points made the best balance between the good of fit and simplicity for 1941-44 cohort by table 3.1. Since the trend of year-specific rate difference for 1941-44 cohort was similar to the trend of 1937-40 cohort, we set the same initial value of breakpoint to fit the model. There was about 55.5% of simulated data that converged to the spline regression model with 2 joint points because for the other 44.5% of the instances, there was only one data point in some segments. Therefore, we expanded the sample size of simulation from 100 to 200 to acquire more converged regression models with required number of segments.

Table 3.5 summarized the statistical results of the fitted spline regression model. Histogram plot of these two breakpoints displayed that their distribution (Figure 3.11) were all asymmetric but unimodal, which indicated that it was more reliable to measure joint points by using median rather than mean value. As the distributions of slopes were roughly symmetric, we used mean value to estimate slope.

Fitted model with the minimum AIC value (-139.00) was displayed in Figure 3.12. The small difference between slope2 and slope3 denoted the difficulty of predicting the value for the second break point. That is why the standard deviation of joint point 2 is so large compared to joint point 1. Like other three cohorts, we could not confirm whether catch-up point occurred or not according to the plot of fitted model (Figure 3.12).

Table 3.5: Summary of statistics for parameters of spline regression by 1941-44 cohort.

Para	Joint point 1	Joint point 2	Slope1	Slope2	Slope3
mean	2.32	7.50	-0.01	-1.68e-04	1.31e-03
Median	2.28	8.00	-0.01	-2.84e-04	1.10e-03
SD	0.224	1.285	0.0019	6.36e-04	1.385e-03
NA	44.5%				

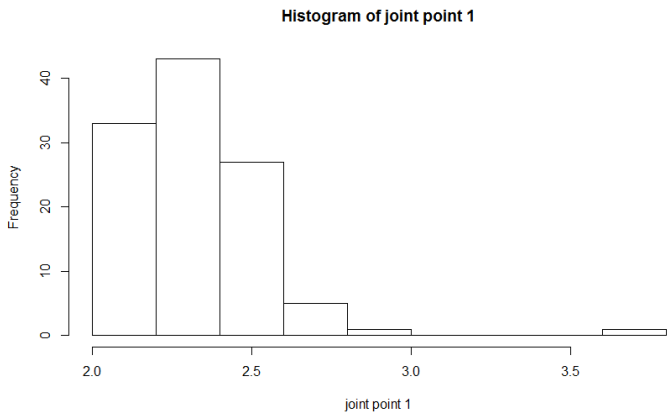
NA: Percentage of cases that are not converged among 200 times simulation.

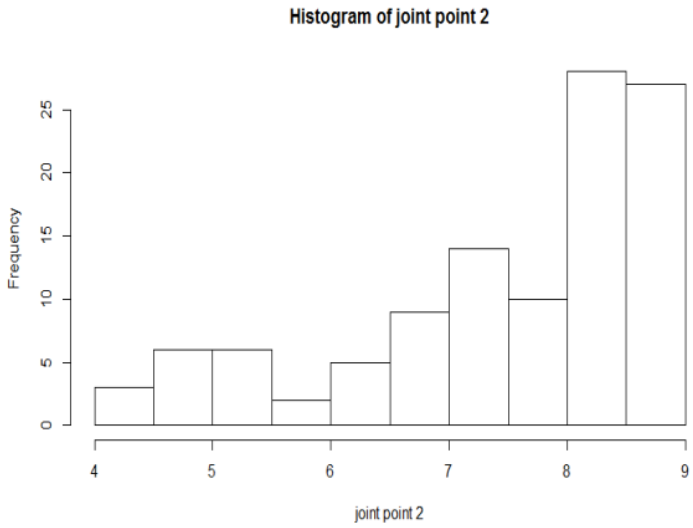
SD: Standard deviation for every parameter we estimated

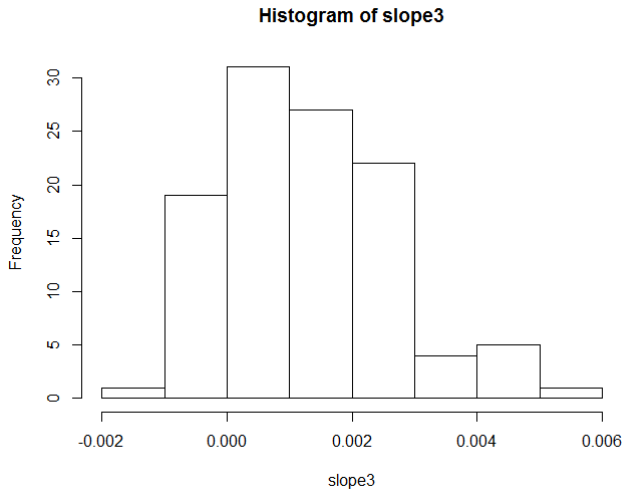
Joint point 1,2: the value of joint points

Slope 1,2, and 3: slopes for the corresponding segments.

Figure 3.11: The histogram plot of breakpoints and slopes for second and third segments of fitted spline regression model after 200 times simulation by 1941-44 cohort. Distributions of all these five parameters are asymmetric.




Histogram of joint point 2
Frequency
0
5
10
15
20
25
4
5
6
7
8
9
joint point 2


Histogram of slope3
Frequency
0
5
10
15
20
25
30
-0.002
0.000
0.002
0.004
0.006
slope3

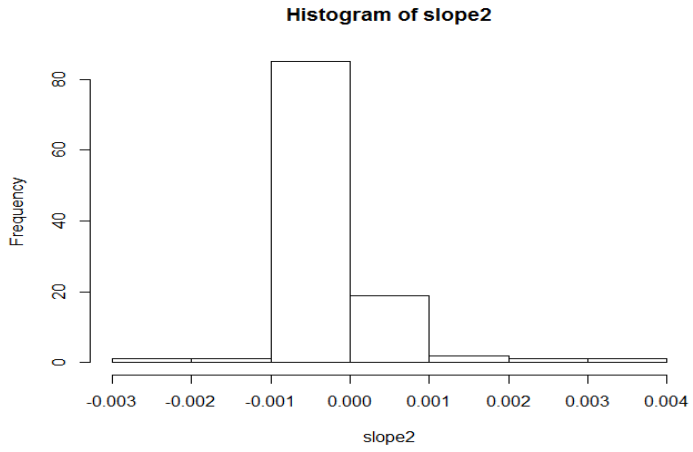
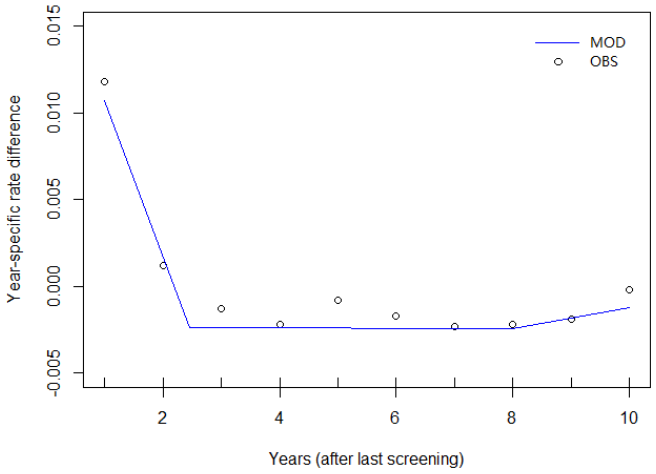

Histogram of slope2
Frequency
0
20
40
60
80
-0.003
-0.002
-0.001
0.000
0.001
0.002
0.003
0.004
slope2

Figure 3.12: Year-specific prostate cancer rate difference for 1941-44 cohort in the Finland of the ERSPC trial as original observed data (OBS), and as predicted by the spline regression model (MOD). The model has the fewest loss information chosen by AIC value.



3.3 Spline regression for artificial data

Follow-up years for all cohorts were not long enough to confirm the place of "catch-up" point we defined. To verify the possibility of finding the "catch-up" point by using spline regression model, inputting function was established to build artificial data for simulation. All slopes and breakpoints especially "catch-up" point were predetermined in inputting function.

In the inputting function, we assumed that "catch-up" point occurred at 18th follow-up years for all cohorts and then added extra 5 years for which year-specific rate difference maintained the value of 0. Since the "catch-up" point in the inputting function occurred after 18 study years. The pattern of the inputting function from 1st study year to 18th year could refer to the estimated

value in our previous simulating analysis combined with the ideal models as Figures 2.3 and 2.4.

According to the different number of breakpoints, inputting function was approximately similar to Figure 2.4 for 1929-32 and 1933-36 cohorts; and to Figure 2.3 for 1937-40 and 1941-44 cohorts, respectively.

After dealing with the inputting function, we needed to acquire the sample variance of rate difference for extra 5 follow-up years. Since we did not have any information about new-added follow-up years, the most straightforward way was to regard the variance of last 5 years of real data as the sample variance of new data.

Then we got the artificial data by simulating the inputting function to follow a normal distribution 100 times. Finally, artificial data was used to fit the spline regression model to find “catch-up” point. By comparing the difference between the estimated value of “catch-up” point and the predetermined value of “catch-up” point, we could assess the reliability of our method to estimate overdiagnosis.

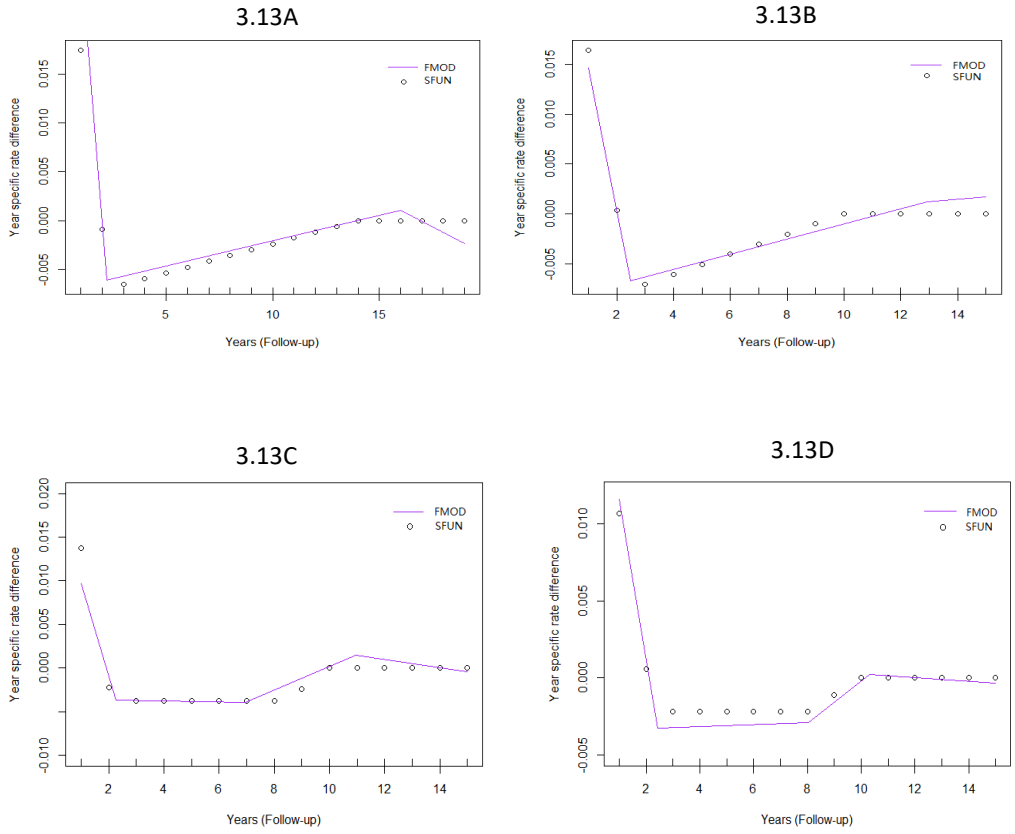
Figure 3.13 showed the inputting function of year-specific rate difference for each cohort, with fitted spline regression model of simulated artificial data painted by the purple line.

For 1929-32 and 1933-36 cohorts, apparent deviation of the third segment (3.13A or 3.13B) between fitted spline regression model and inputting function

could be explained by the significant sample variance we set for artificial data, which was almost as 1.5 times large as other two cohorts.

In each predetermined function, the “catch-up” point was determined to occur in 18th years. Since there are only two screenings for 1929-32 cohort, “catch-up” point was assumed to occur in 14th follow-up year after last screening time. Otherwise, catch-up point was set at 10th follow-up year after the last screening round.

Figure 3.13: Inputting function(SFUN) and as fitted spline regression model (FMOD) after simulating the inputting function.3.13A, 3.13B, 3.13C and 3.13D were corresponding to 1929-32, 1933-36, 1937-40, and 1941-44 cohort, respectively.



For 1929-32 cohort, estimated slope of the last segment was -1.54e-04 which was roughly equal to 0. So, we could confirm that the second joint point was exactly the “catch-up” point. Moreover, the histogram plot of parameter joint point 2 (catch-up point) displayed that the distribution is asymmetric, which implied that median value was more precise to estimate the place of joint point 2 than the mean value. So, the error of the estimation for “catch-up” point would be (14-13.07)/14*100%=6.64%

Table 3.6: Summary of statistics for parameters of spline regression by 1929-32 cohort.

Para	Joint point 1	Joint point 2	Slope1	Slope2	Slope3
mean	2.36	12.61	-1.84e-02	6.67e-04	-1.54e-04
Median	2.35	13.07	-1.79e-02	6.44e-04	1.10e-03
True	2.30	14.00	-0.018	5.95e-04	0
SD	0.195	3.945	3.156e-03	5.20e-04	-5.684e-05
NA	20%				

NA: Percentage of cases that are not converged among 100 times simulation.

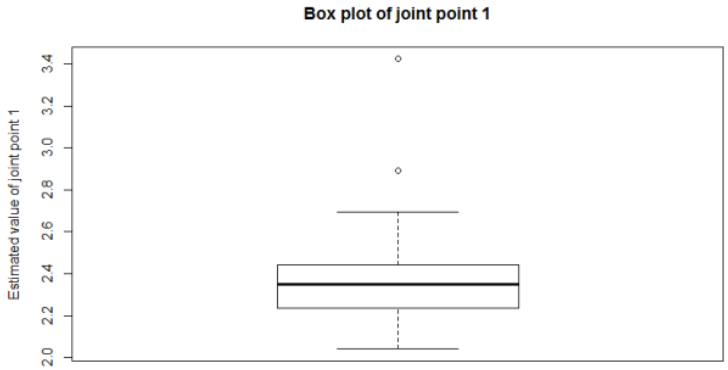
SD: Standard deviation for every parameter we estimated

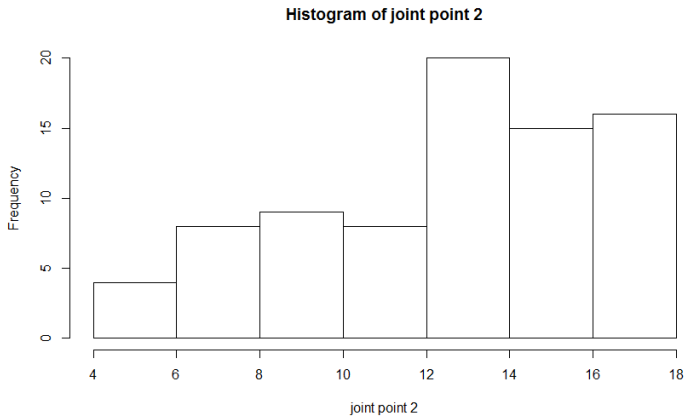
Joint point 1,2: the value of joint points

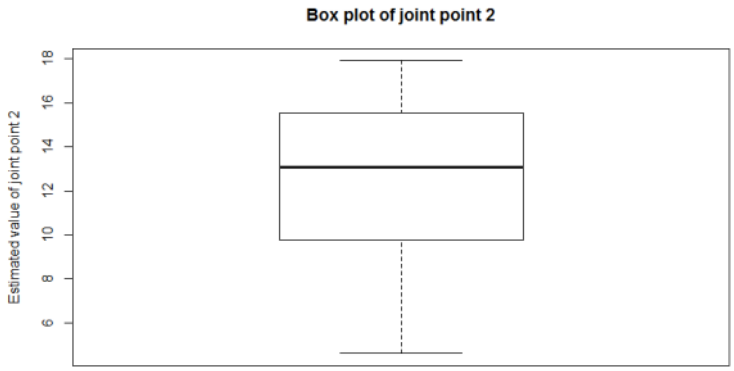
Slope 1,2, and 3: slopes for the corresponding segments.

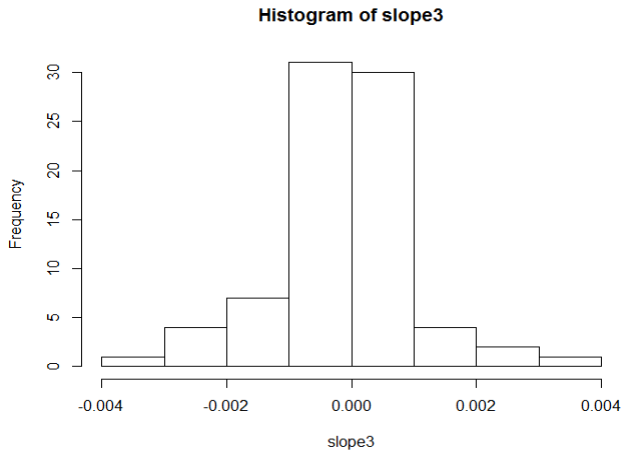
True Value: the value we predetermined for the inputting function

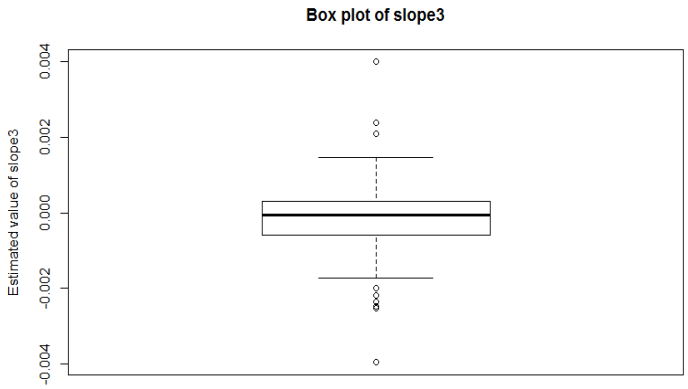
Figure 3.14: The histogram and box plots of breakpoints and slopes of fitted spline regression model after 100 times simulation for artificial data by 1929-32 cohort. Distribution of slope for the last segment was symmetric, and 0 was contained in its whisker box range.











For 1933-36 cohort, asymmetric distribution of slope3 indicated that median value (-0.0009407) is a reasonable choice to estimate this parameter. As the estimated value for the slope of the last segment was almost 0, we could infer that Joint point 2 was the “catch-up” point, and its estimated result was 12.91 since the histogram plot of joint point 2 showed its distribution was asymmetric. So, the error of estimation for “catch-up” point would be (12.91-10)/10*100%=29.1%

Table 3.7: Summary of statistics for parameters of spline regression by 1933-36 cohort.

Para	Joint point 1	Joint point 2	Slope1	Slope2	Slope3
mean	2.58	11.71	-1.63e-02	1.30e-03	-1.48e-03
Median	2.53	12.91	-1.64e-02	1.16e-03	-9.41e-04
True	2.41	10	-1.61e-02	1.01e-03	0
SD	0.264	2.739	3.613e-03	6.651e-04	1.925e-03
NA	39%				

NA: Percentage of cases that are not converged among 100 times simulation.

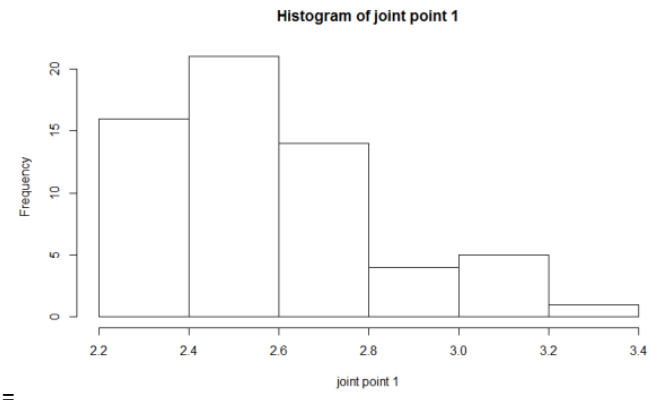
SD: Standard deviation for every parameter we estimated

Joint point 1,2: the value of joint points

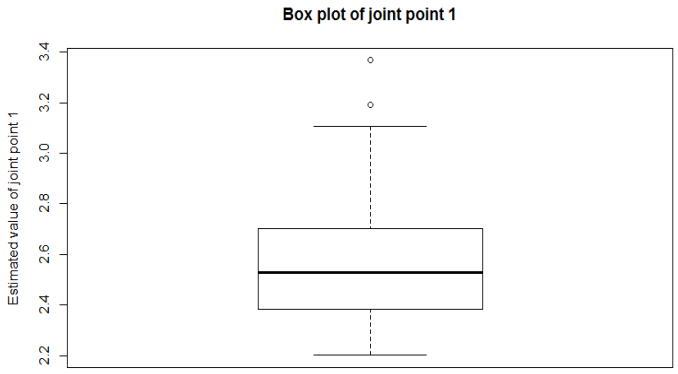
Slope 1,2, and 3: slopes for the corresponding segments.

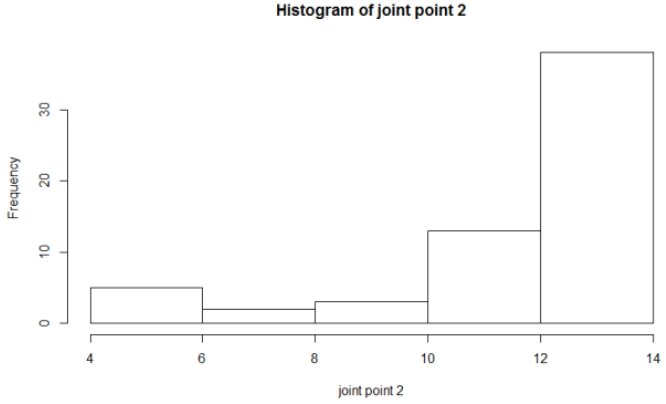
True Value: the value we predetermined for the inputting function

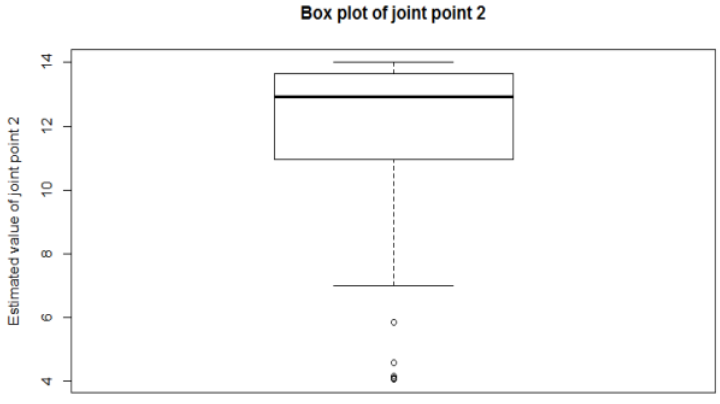
Figure 3.15: The histogram and box plots of breakpoints and slopes of fitted spline regression model after 100 times simulation for artificial data by 1933-36 cohort. In the box plot of slope3, 0 was contained in the whisker range.

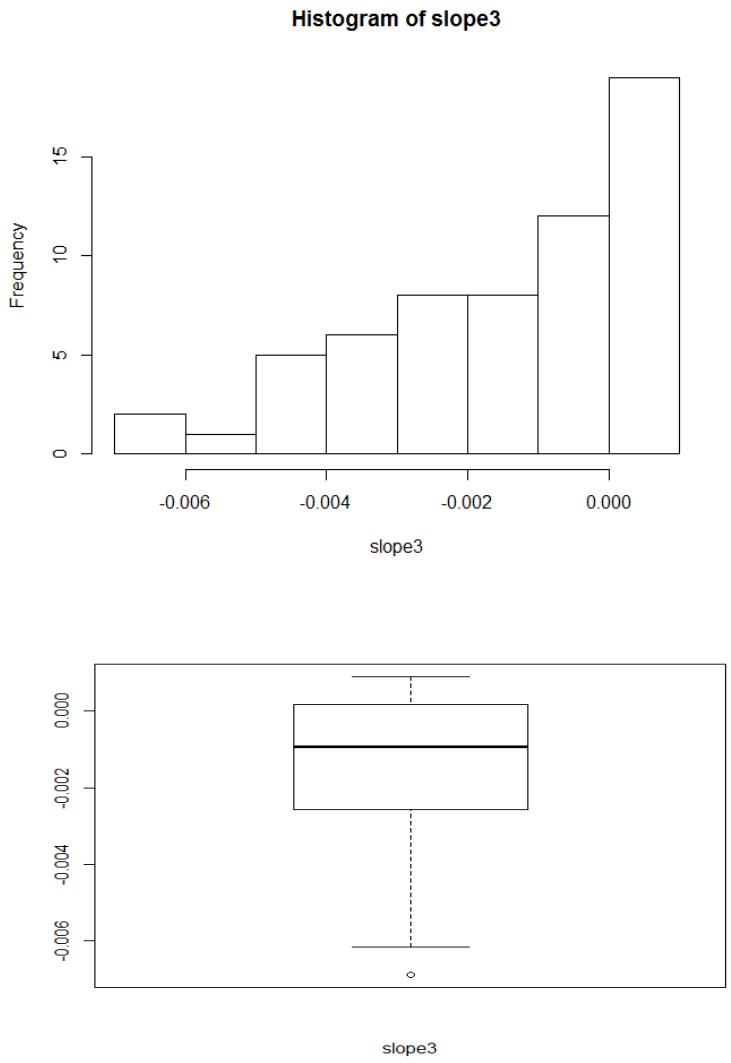


=


Box plot of joint point 1
Estimated value of joint point 1
2.2
2.4
2.6
2.8
3.0
3.2
3.4


Histogram of joint point 2
Frequency
0
10
20
30
4
6
8
10
12
14
joint point 2


Box plot of joint point 2
Estimated value of joint point 2
4
6
8
10
12
14



For the 1937-40 cohort, unlike the former two cohorts, there were in total four segments in the inputting function and fitted model (as Figure 3.13C). Not considering the outliers, distribution of slope4 was primarily symmetric (as Figure 20). As the estimated value (-1.95e-04) of slope3 was almost 0, it could be believed that joint point 3 was the “catch-up” point, and its estimated result was 10.15 (median value) since the histogram plot of joint point 3 showed that its distribution was asymmetric. So, the error of estimation for “catch-up” point was 1.5%

Table 3.8: Summary of statistics for parameters of spline regression by the 1937-40 cohort.

para	Joint point 1	Joint point 2	Joint point 3
mean	2.13	7.76	10.44
Median	2.10	8	10.15
True value	2.10	8.42	10
sd	0.135	1.234	1.233
NA	60%		

para	Slope1	Slope2	Slope3	Slope4
mean	-1.54e-02	-1.14e-04	2.12e-03	-1.95e-04
Median	-1.55e-02	-1.66e-04	2.02e-03	-1.59e-04
True value	-1.60e-02	0	2.41e-03	0
sd	2.288e-03	5.121e-04	1.129e-03	5.861e-04
NA	60%			

NA: Percentage of cases that are not converged among 100 times simulation.

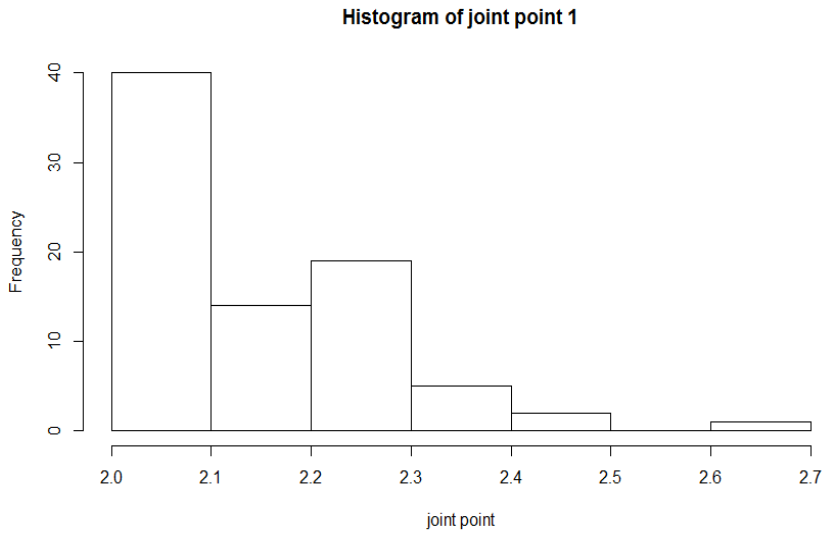
SD: Standard deviation for every parameter we estimated

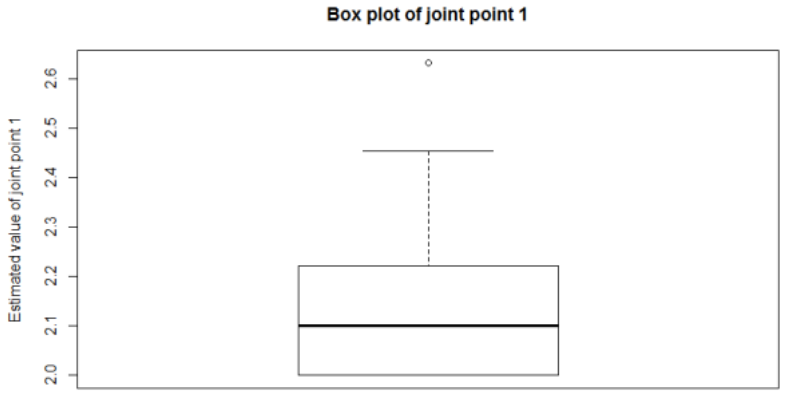
Joint point 1,2: the value of joint points

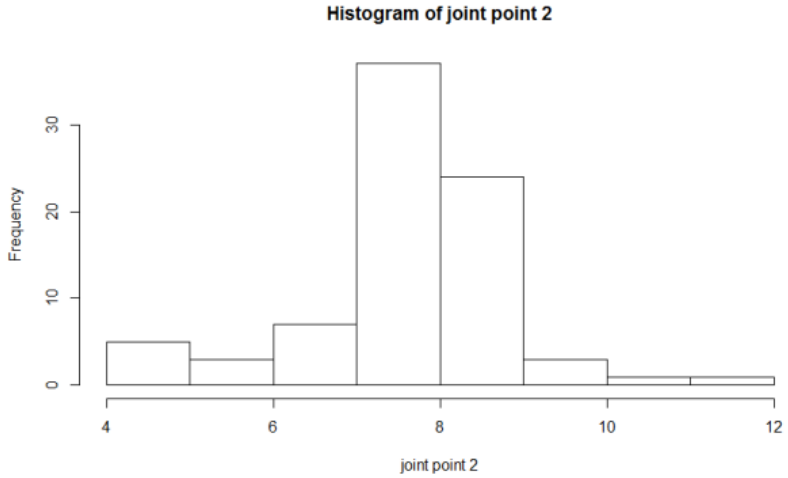
Slope 1,2, 3 and 4: slopes for the corresponding segment.

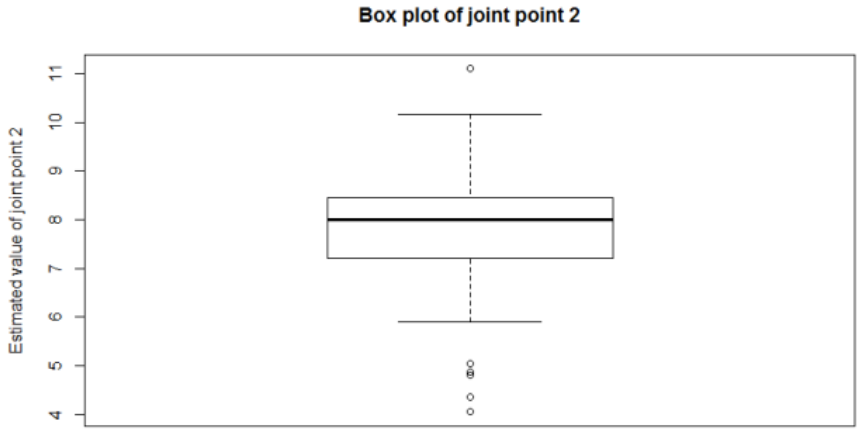
True Value: the value we predetermined for the inputting function

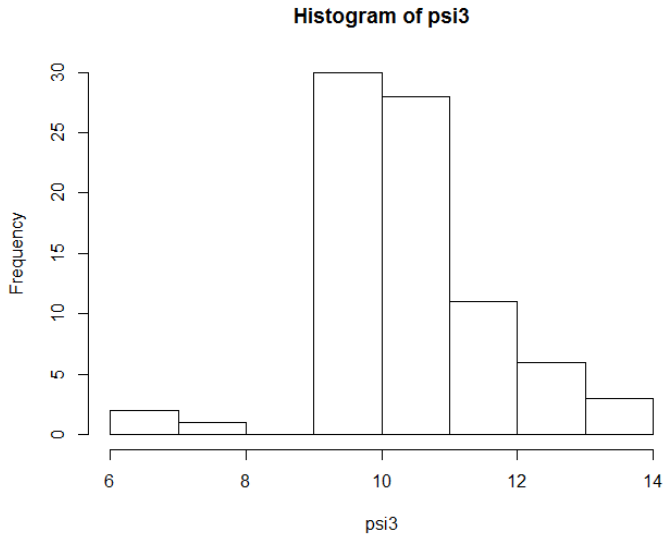
Figure 3.16: The histogram and the box plot of breakpoints and slopes of fitted spline regression model after 100 times simulation for artificial data by the 1937-40 cohort.

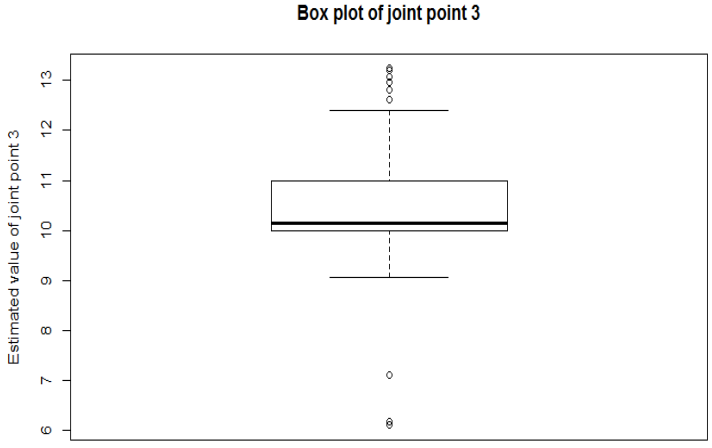

Histogram of joint point 1
Frequency
0
10
20
30
40
2.0
2.1
2.2
2.3
2.4
2.5
2.6
2.7
joint point

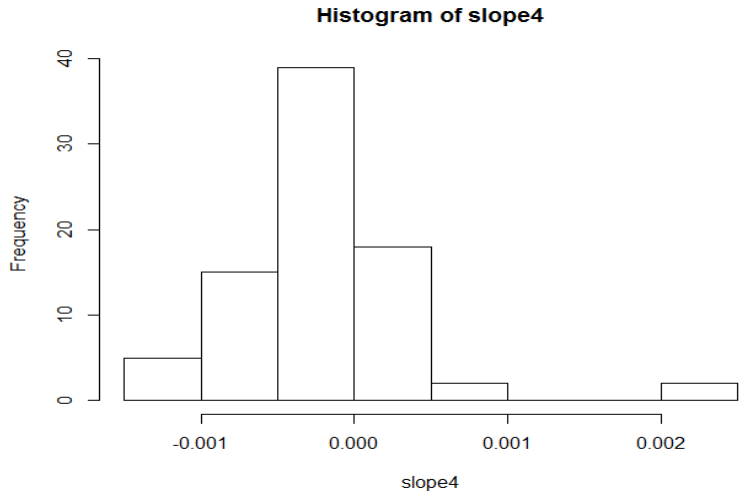

Box plot of joint point 1
Estimated value of joint point 1
2.0
2.1
2.2
2.3
2.4
2.5
2.6

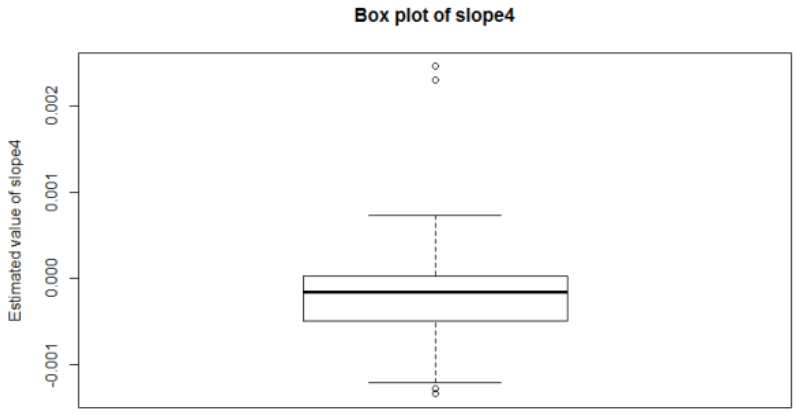

Histogram of joint point 2
Frequency
0
10
20
30
4
6
8
10
12
joint point 2


Box plot of joint point 2
Estimated value of joint point 2
4
5
6
7
8
9
10
11


Histogram of psi3
Frequency
0
5
10
15
20
25
30
6
8
10
12
14
psi3


Box plot of joint point 3
Estimated value of joint point 3
6
7
8
9
10
11
12
13





In the box plot of slope4, 0 was contained in the whisker range.

For the 1941-44 cohort, there were also four segments in the inputting function and fitted model (as Figure 3.13D). Distribution of slope4 was basically symmetric. As the estimated value (-1.90e-04) for the slope of the last segment was almost 0, it seemed that Joint point 3 was the “catch-up” point, and its estimated result was 10.69(median value) since the histogram plot of joint point 3 showed its distribution was asymmetric. So, the error of estimation for “catch-up” point would be 6.9%.

Table 3.9: Summary of statistics for parameters of spline regression by the 1941-44 cohort.

para	Joint point 1	Joint point 2	Joint point 3
mean	2.33	7.058	11.01
Median	2.29	7.11	10.69
True value	2.28	8.00	10
sd	0.237	1.563	1.725
NA	60%		

para	Slope1	Slope2	Slope3	Slope4
mean	-9.92e-03	-2.22e-04	9.77e-4	-1.90e-04
Median	-1.01e-02	-1.57e-04	8.64e-04	-1.29e-04
True value	-1.00e-02	0	1.10e-03	0
sd	2.032e-03	6.805e-04	7.54e-04	9.36e-04
NA	60%			

NA: Percentage of cases that are not converged among 100 times simulation.

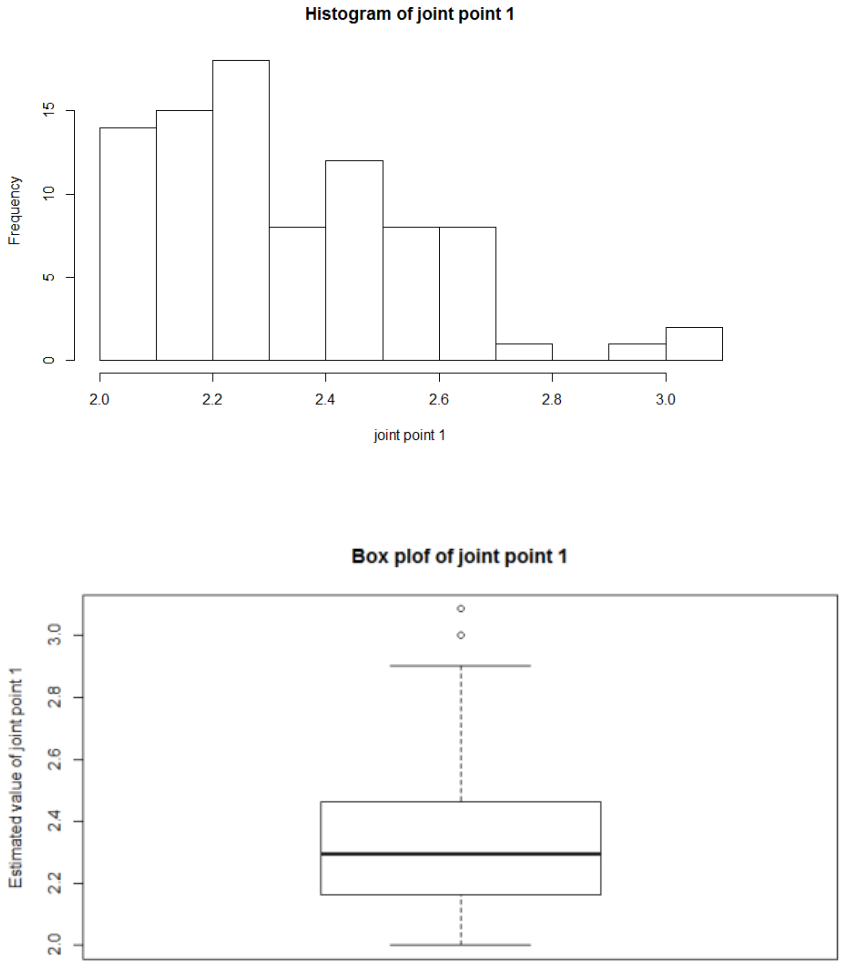
SD: Standard deviation for every parameter we estimated

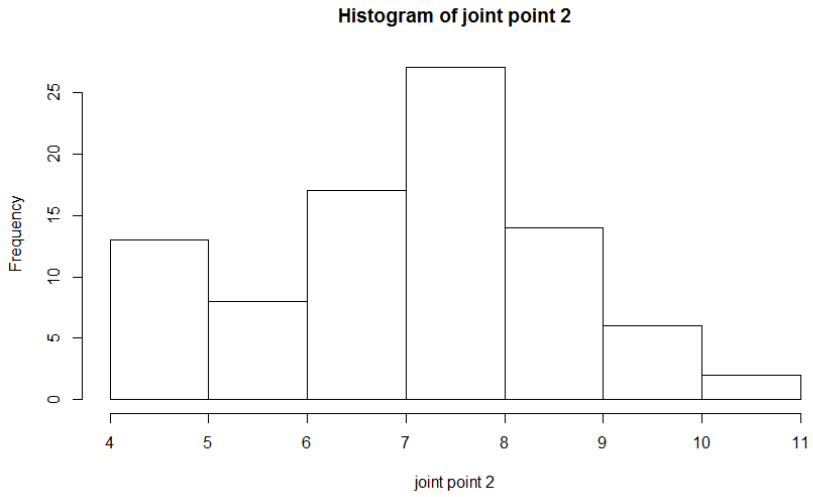
Joint point 1,2: the value of joint points

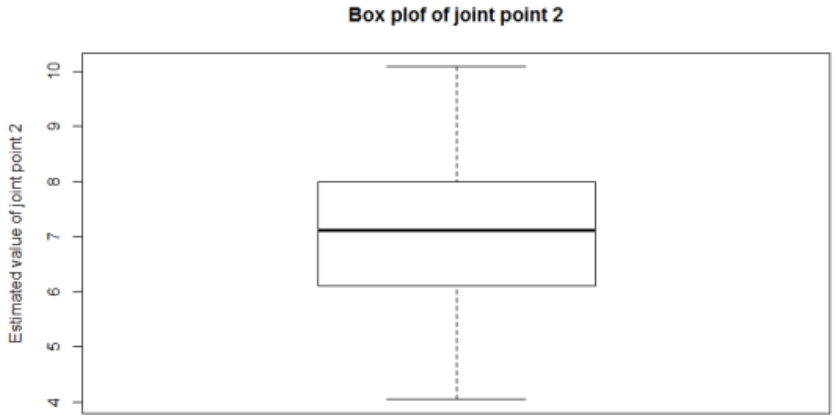
Slope 1,2, 3 and 4: slopes for the corresponding segment.

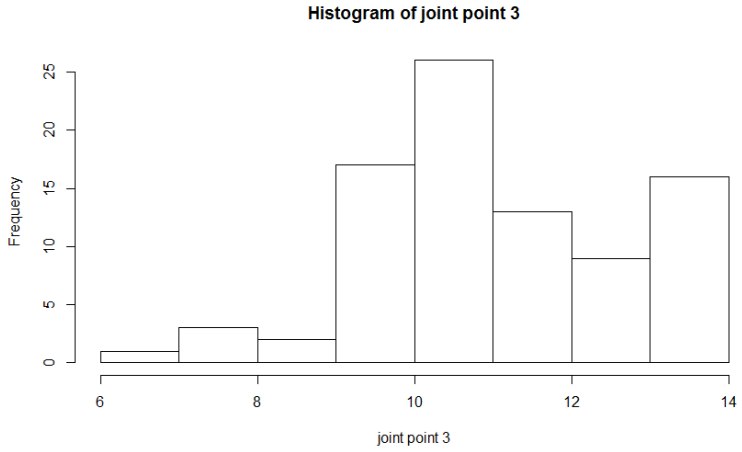
True Value: the value we predetermined for the inputting function

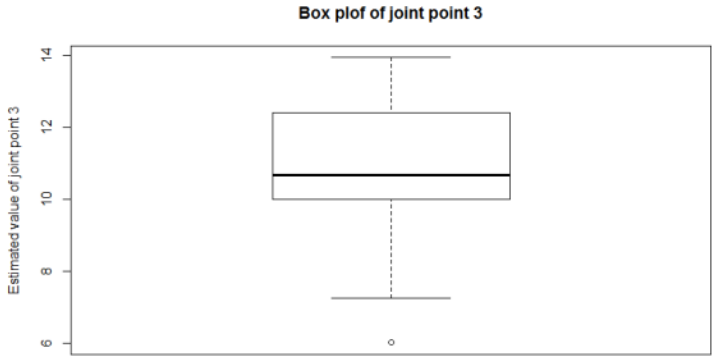
Figure 3.17: The histogram and the box plots of breakpoints and slopes of fitted spline regression model after 100 times simulation for artificial data by cohort 1941-44. In the box plot of slope4, 0 was contained in the whisker range.

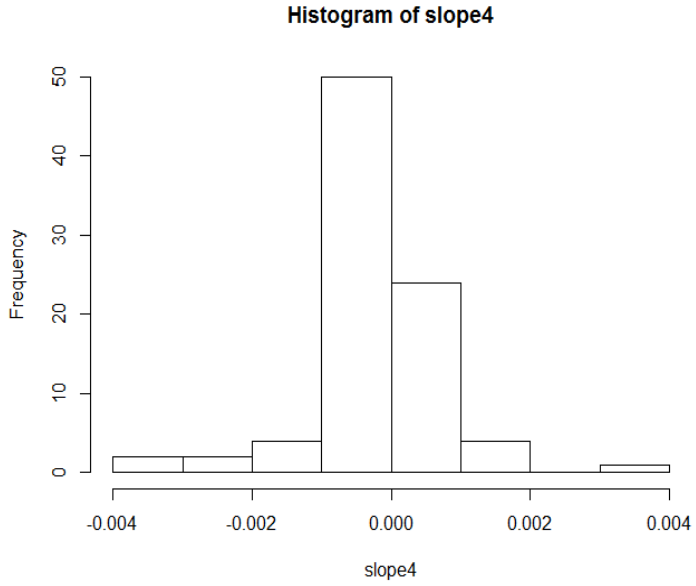


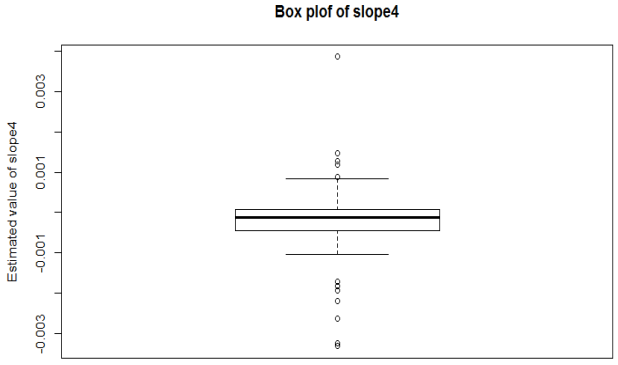

Histogram of joint point 2
Frequency
0
5
10
15
20
25
4
5
6
7
8
9
10
11
joint point 2


Box plof of joint point 2
Estimated value of joint point 2
4
5
6
7
8
9
10


Histogram of joint point 3
Frequency
0
5
10
15
20
25
6
8
10
12
14
joint point 3


Box plof of joint point 3
Estimated value of joint point 3
6
8
10
12
14

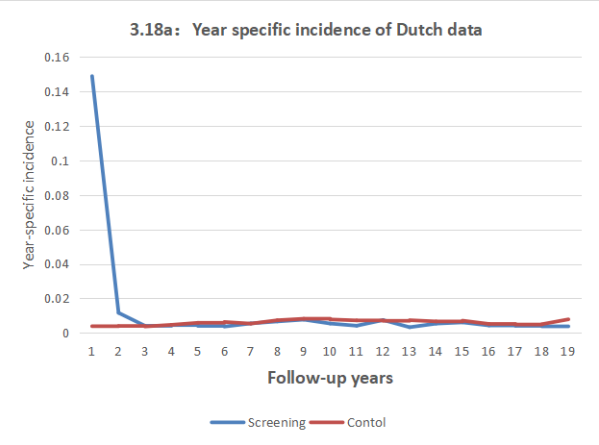

Histogram of slope4
Frequency
0
10
20
30
40
50
-0.004
-0.002
0.000
0.002
0.004
slope4


Box plof of slope4
Estimated value of slope4
-0.003
-0.001
0.001
0.003

3.4 Spline regression model for Netherlands section

Excess-incidence approach was also tried to apply to the Netherlands data. For the Netherlands section, men in the age group 54–74 years were randomized to either a screening group or a control group. Men randomized to the control group were not offered PSA testing. Men in the screening group were invited for PSA testing. A PSA level 3.0 ng/ml was used as the indication for biopsy. Men were invited for next screening every 4 years if no PCa was detected at previous screening or during the preceding interval.

Figure 3.18: Year-specific rate difference of the Netherlands data between the control arm and screening group who only received screening once only.



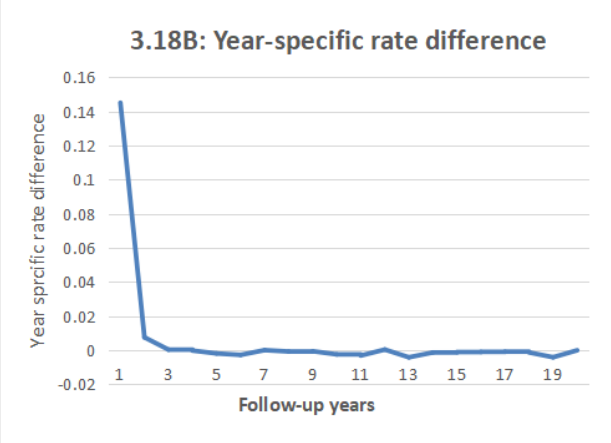


Figure 3.18 showed the pattern of the year specific-rate difference for the Netherlands data. The annual incidence acquired its highest value at the very first screening round, then dropped sharply to 0 and finally maintained this level until the end of follow-up years.

Figure 3.19 plotted the best model with the minimum AIC value (-199.5264). There were two segments of year-specific rate difference in this model. Rate difference dropped sharply to the bottom after 2 follow-up years (after the last screening time), then almost leveled-off all the time. It seemed the “catch-up” point was the joint point. The estimate value of slope2 also verified this. According to our definition of “catch-up” point, the year-specific incidence became stable once the screening stopped. However, this trend of year-specific rate in this screening group was unable to follow the normal pattern of annual incidence in a screening group for a randomized trial.

Table 3.10: Summary of statistics for parameters of spline regression for the Netherland data

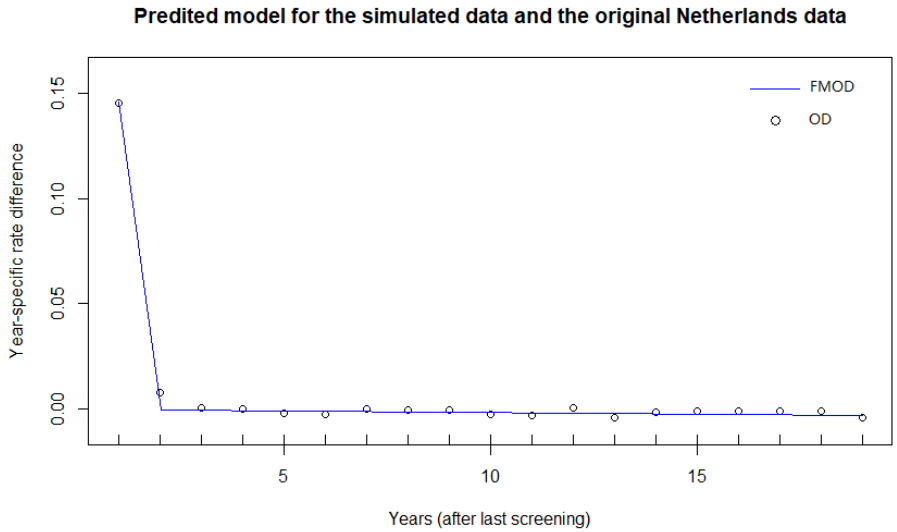
Para	Joint point	Slope1	Slope2
mean	2.059	-0.1373	-8.134e-05
SD	0.0123	0.0046	8.6114e-05
NA	0%		

NA: Percentage of cases that are not converged among 100 times simulation.

SD: Standard deviation for every parameter we estimated

Joint point 1: the value of joint points

Figure 3.19: Observed year-specific incidence difference of Netherlands (OD) and as fitted spline regression model (FMOD) after simulating the Netherlands data.



Chapter 4

Discussion

In the method section, we verified the reliability of this spline regression model to confirm the value of “catch-up” point. The error of the estimation of “catch-up” point was 6.24%, 29.1%, 1.5%, and 6.9% for these different four cohorts, respectively. The acceptable error for the spline regression mode indicated that sufficient follow-up years as long as the longest preclinical period for screening stabilizing should be provided to acquire an unbiased empirical estimate of the case of overdiagnosis. Based on our results, we could conclude that current 8 or 10 follow-up years were not long enough for us to confirm whether the incidence stabilized or not. The conclusion, it took as long as 10-14 years for the change of the incidence caused by PSA screening becoming stable (Auvinen A, 2018) acquired from a study about the estimate of lead-time in prostate cancer verified our result. Therefore, it was believed that the difference of cumulative incidence for prostate cancer between the screening group and control group still tended to narrow. Under this circumstance, the estimation of overdiagnosis could be calculated from the difference of cumulative incidence in the last year after screening stopped, nevertheless it would be overestimated.

Regarding the data from Netherlands section, the different pattern of year-specific rate indicated it was not available for the Netherlands data to use spline regression model to find “catch-up” point due to the extremely high

incidence of prostate cancer during their first screening time. This phenomenon could be partly explained by the fact that the cut-off value (3 ng/nl) used as biopsy indication in the Netherland center was lower than cut-off value (4 ng/nl) in the Finland data. Therefore, more people would be detected as prostate cancer during their screening. This unusual high incidence in the first point pattern (Figure 22B) made the slope for the second segment almost equal to 0 when fitting the Dutch data after simulation to spline regression model. The result may lead to a wrong conclusion that the "catch-up" point occurred after the first screening. Then the overdiagnosis rate calculated by the cumulative rate difference according to the catch-up point would be overestimated.

Table 4.1 verified our assumption that the excess incidence still tended to decrease for the catch-up point did not occur yet. Under this circumstance, the cumulative incidence difference in the current longest follow-up year was the best estimation of overdiagnosis frequency. Therefore, the best estimation of prostate cancer overdiagnosis frequency for men who were born in 1929-32 was 0.004. Corresponding standard error for cumulative incidence in screened group and control group were 0.0061 and 0.005. By equation (4), the confidence interval was (-0.011,0.019). The corresponding overdiagnosis rate was $\frac{0.004}{0.176} \times 100\% = 2.27\%$ by equation (3).

In theory, overdiagnosis rate increased with age because of the combined effect of a higher detection rate and of a higher mortality rate resulting from other causes for elder people (Zappa, Ciatto, Bonardi & Mazzotta, 1998).

However, the irregularly small overdiagnosis rate (2.27%) for the oldest men could be explained by the fact that there were only two screening rounds and longer follow-up years for the 1929-32 cohort.

Table 4.1: Incidence excess and estimate of overdiagnosis by birth cohorts.

AGE AT THE START OF SERVICE SCREENING	YEARS AFTER SCREENING STOPPED	NO. OF OVERDIAGNOSIS (95%CI)	OVERDIAGNOSIS RATE (%)
67-70	10	0.005(-0.010,0.020)	3.03
	11	0.002(-0.012,0.016)	1.18
	12	0.004 (-0.011,0.019)	2.27
63-66	7	0.021(9.12e-03,0.033)	13.4
	8	0.026(0.013,0.039)	15.7
	9	0.026(0.013,0.039)	15.4
59-62	7	0.016(6.51e-03,0.025)	13.2
	8	0.013(2.96e-03,0.023)	10.2
	9	0.015(4.38e-03,0.026)	11.4
56-58	7	0.0122(5.07e-03,0.019)	14.0
	8	0.0100(2.47e-03,0.17)	11.1
	9	0.0099(1.67e-03,0.018)	10.2

Table 4.2 summarized the estimation of overdiagnosis in other studies. The table suggested that 2.9–88.1% could be regarded as an overdetection. Such a substantial variation was led by the fact that estimates of overdiagnosis are generally presented as a ratio, with the numerator being the estimated number of cases overdiagnosed and with many options for the denominator (Etzioni, R. et al., 2013). Studies conducted by Etzioni, Telesca and us reported overdiagnosis as a fraction of screening-detected cases. Others presented the number overdiagnosed as a fraction of the total number of cases detected, or the total number invited to screening. In addition to the definition, the different

methodologic approach also attributed to the wide range of the estimation. We made the point that each method had its limitations.

In most modeling studies, investigators used disease incidence under screening to make the distribution of the lead time or the natural history of the disease and estimated the corresponding frequency of overdiagnosis. The strength of this method was not constrained by time and resource, whereas the limitations were the lack of transparency and the difficulty in evaluating a model like a black box critically.

Unlike modeling studies, excess of incidence approach provides a direct estimate of overdiagnosis. For excess incidence studies, there are two challenges for investigators; one is requirement of sufficient follow-up years, the other is how to observe the incidence without screening. Investigators utilized a different method to impute incidence data without screening. Zappa and Ciatto calculated incidence data without screening from prescreening trend, while the data in Schröder studies were taken from the randomized clinical trial. Sufficient follow-up years may seem to resolve the timing problem, but long-term studies mean it is more challenging to stop men in the control arm from being screened by PSA testing during the study years. Therefore, contamination rate of control group should be considered to calculate final overdiagnosis rate, especially for our case, the Finland data. A report about contamination rate in ERSPC studies showed that 10% of men in the screening group had a PSA testing before their first screening of the trial. More recently, it has been estimated that 50% of the men in the control arm in Finland have been tested at least once in the first eight years of follow-up

(Nevalainen, J et al.,2017). Such a high contamination rate in the control arm was bound to reduce the excess incidence between the two groups and led to an underestimated result of overdiagnosis rate even the follow-up years were long enough for the incidence becoming stable.

Table 4.2: Summary of 6 modeling studies and 3 excess-incidence studies quantifying overdiagnosis rate from PSA testing.

Approach	Researcher	Study Years	Data	Estimation of overdiagnosis
Modeling Study	Draisma, 2003	2003	ERSPC Rotterdam	48%
	Etzioni, 2002	1988-1998	U.S. SEER9	29% in white persons, 44% in black persons
	Roman Gulati,2014	1975-2005	U.S. SEER9	2.9-88.1% depending on age, Gleason score, and PSA level
	Telesca 2007	1975–2000	U.S. SEER9	22.7% in white persons, 34.4% in black persons
	Wu 2012	1996-2005	ERSPC Finland	3.4%
Excess Incidence	Zappa 1998	1992-1995	Italy	51% for constant incidence; 25% for 2% annual incremental incidence
	Schröder et al., 2009	1991-2006	ERSPC	48 cases of 1410 screened men
	Ciatto.S 2006	1991-1994	Italy	66%

U.S. SEER9：core 9 catchment areas of the Surveillance, Epidemiology, and End Results program.

In conclusion, we proved the feasibility of spline regression model to find the “catch-up” point in which incidence become stable. Given sufficient follow-up years, we could calculate how long it takes to elapse before the excess cumulative incidence calculation would produce an unbiased estimation. Based on current data, the cumulative difference in the longest follow-up year overestimated but was close to the overdiagnosis rate under the assumption there was no contamination in the control arm. Not considering the contamination rate, the estimate of overdiagnosis rate was 2.27%,15.4%, 11.4%, and 10.2% for 1929-32, 1933-36,1937-40, and 1941-44 cohorts, respectively.

References

Auvinen A, e. (2018). *Lead-time in prostate cancer screening (Finland). - PubMed - NCBI*. *Ncbi.nlm.nih.gov*. Retrieved 13 January 2018, from https://www.ncbi.nlm.nih.gov/pubmed/12020110

Bell, K., Del Mar, C., Wright, G., Dickinson, J., & Glasziou, P. (2015). Prevalence of incidental prostate cancer: A systematic review of autopsy studies. *International Journal Of Cancer*, *137*(7), 1749-1757.

Biesheuvel, C., Barratt, A., Howard, K., Houssami, N., & Irwig, L. (2007). Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. *The Lancet Oncology*, *8*(12), 1129-1138.

Cancer Facts & Figures 2017. (2018). *Cancer.org*. Retrieved 13 January 2018, from https://www.cancer.org/research/cancer-facts-statistics/all-cancer-facts-figures/cancer-facts-figures-2017.html

Carter, J., Coletti, R., & Harris, R. (2015). Quantifying and monitoring overdiagnosis in cancer screening: a systematic review of methods. *BMJ*, *350*(jan07 5), g7773-g7773.

Ciatto, S., et al. “Contamination by Opportunistic Screening in the European Randomized Study of Prostate Cancer Screening.” *BJU International*, Wiley/Blackwell (10.1111), 13 Jan. 2004

Ciatto, S., Gervasi, G., Bonardi, R., Frullini, P., Zendron, P., Lombardi, C., . . . Zappa, M. (2005). Determining overdiagnosis by screening with DRE/TRUS or PSA (Florence pilot studies, 1991–1994). *European Journal of Cancer, 41*(3), 411-415.

Draisma, G., & de Koning, H. (2003). MISCAN: estimating lead-time and over-detection by simulation. *BJU International*, *92*(s2), 106-111.

Draisma, G., Etzioni, R., Tsodikov, A., Mariotto, A., Wever, E., & Gulati, R. et al. (2009). Lead Time and Overdiagnosis in Prostate-Specific Antigen Screening: Importance of Methods and Context. *JNCI Journal Of The National Cancer Institute*, *101*(6), 374-383.

Epidemiology of Prostate Cancer | *touchONCOLOGY*.
(2018). *Touchoncology.com*. Retrieved 13 January 2018, from http://www.touchoncology.com/articles/epidemiology-prostate-cancer

Etzioni, R. (2002). Overdiagnosis Due to Prostate-Specific Antigen Screening: Lessons From U.S. Prostate Cancer Incidence Trends. *Cancerspectrum Knowledge Environment*, *94*(13), 981-990.

Etzioni, R., Cha, R., Feuer, E., & Davidov, O. (1999). Asymptomatic Incidence and Duration of Prostate Cancer. *The Journal of Urology*, 162(1), 265-266.

Etzioni, R., Gulati, R., Mallinger, L., & Mandelblatt, J. (2013). Influence of Study Features and Methods on Overdiagnosis Estimates in Breast and Prostate Cancer Screening. *Annals of Internal Medicine,158*(11), 831.

Etzioni, R., Tsodikov, A., Mariotto, A., Szabo, A., Falcon, S., & Wegelin, J. et al. (2007). Quantifying the role of PSA screening in the US prostate cancer mortality decline. *Cancer Causes & Control*, *19*(2), 175-181.

Gulati, R., Inoue, L., Gore, J., Katcher, J., & Etzioni, R. (2014). Individualized Estimates of Overdiagnosis in Screen-Detected Prostate Cancer. *JNCI Journal Of The National Cancer Institute*, *106*(2), djt367-djt367.

Hakama M, Auvinen A. Cancer screening. In: Heggenhougen K, Quah SR, eds. International *Encyclopedia of Public Health. San Diego, CA: Academic Press, 2008*:464-80.

Nevalainen, J., Stenman, U., Tammela, T. L., Roobol, M., Carlsson, S., Talala, K., . . . Auvinen, A. (2017). What explains the differences between centers in the European screening trial? A simulation study*. Cancer Epidemiology, 46,* 14-19.

Paci, E., & Duffy, S. (2005). Overdiagnosis and overtreatment of breast cancer: Overdiagnosis and overtreatment in service screening. *Breast Cancer Research*, *7*(6).

Pashayan, N., Duffy, S., Pharoah, P., Greenberg, D., Donovan, J., & Martin, R. et al. (2009). Mean sojourn time, overdiagnosis and reduction in advanced stage prostate cancer due to screening with PSA: implications of sojourn time on screening. *British Journal of Cancer*, *100*(7), 1198-1204.

Schröder, F., Hugosson, J., Roobol, M., Tammela, T., Ciatto, S., & Nelen, V. et al. (2009). Screening and Prostate-Cancer Mortality in a Randomized European Study*. New England Journal of Medicine,* 360(13), 1320-1328*.*

Smith RA, e. (2018). *American Cancer Society guidelines for the early detection of cancer, 2003. - PubMed - NCBI*. *Ncbi.nlm.nih.gov*. Retrieved 13 January 2018, from https://www.ncbi.nlm.nih.gov/pubmed/12568442

Tsodikov, A., Szabo, A., & Wegelin, J. (2006). A population model of prostate cancer incidence. *Statistics In Medicine*, *25*(16), 2846-2866.

Telesca, D., Etzioni, R., & Gulati, R. (2007). Estimating Lead Time and Overdiagnosis Associated with PSA Screening from Prostate Cancer Incidence Trends. *Biometrics*, *64*(1), 10-19.

Wu GHM, Auvinen A, Maattanen L, Tammela TLJ, Stenman UH, Hakama M, et al. Number of screens for overdetection as an indicator of absolute risk of overdiagnosis in prostate cancer screening. *Int J Cancer* 2012; 131:1367-75.

Zappa, M., Ciatto, S., Bonardi, R., & Mazzotta, A. (1998). Overdiagnosis of prostate carcinoma by screening: An estimate based on the results of the Florence Screening Pilot Study. *Annals of Oncology*, *9*(12), 1297-1300.