ABSTRACT

Title of Thesis: A STUDY OF THE EFFICACY OF SCREENING FOR BREAST CANCER BASED ON THE HEALTH INSURANCE PLAN DATA

Lina Kurniasih Himawan, Master of Arts, 2004

Thesis directed by: Dr. Grace L. Yang
Department of Mathematics

In this thesis we would like to assess the benefit of screening in breast cancer detection. For this purpose, the Health Insurance Plan data will be analyzed.

If screening is beneficial, we expect to see higher early stage detection in the study group compare to the control group which hopefully will lead to a better survival probability.

Our analysis shows that there is a significantly higher combined stage I detection proportion in the study group compared to the control group (50.00% in the control group as opposed to 67.77% in the study group). Furthermore, the study group has a significantly lower proportion of breast cancer deaths than the control group. The Kaplan-Meier estimate shows that the study group has a

better survival function than the control group. In addition to these, we shows that there exists no significant lead time. Lastly, we construct a Markov chain model for future analysis.

A STUDY OF THE EFFICACY OF SCREENING FOR BREAST CANCER BASED ON THE HEALTH INSURANCE PLAN DATA

by

Lina Kurniasih Himawan

Thesis submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Master of Arts
2004

Advisory Committee:

Dr. Grace L. Yang, Chair
Dr. Paul J. Smith
Dr. Ryszard Syski

DEDICATION

To Angga who always believes in me even when I lost my faith in me

And to Papi and Mami for their never ending support

ACKNOWLEDGEMENTS

On the completion of my graduate study, I would like to give thanks to Heavenly Father who has given me strength and patience to finish my thesis. It has been a long journey, but I am grateful for His guidance and blessings along the way.

I would like to thank my advisor Dr. Grace L. Yang for her ceaseless guidance and patience in the completion of my thesis. Her vast experience is a priceless treasure and I am really fortunate and honored to have the opportunity to learn from her.

I would also like to thank Dr. Philip Prorok, Dr. Jian-Lun Xu and Diane Erwin of the National Institute of Health for their willingness to provide the Health Insurance Plan data and their assistance in understanding the data.

Finally, I would like to thank my parents, my sister, my brother, Angga and my friends, the Sultanas family, Bettie and Nuk, for their constant and continuous support and encouragement during my study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Assessing the efficacy of screening for cancer is an outstanding problem in cancer research. The problem has been studied by medical researchers around the world. In a screening trial for breast cancer the participating women are divided into a study group and a control group according to a certain randomization procedure. Women in the study group will be offered a number of screening examinations for breast cancer over a period of time while women in the control group will not be invited for any screening. The participants in both groups are followed up for a long period of time, years beyond the screening period, to record their incidence and mortality of breast cancer. The principal question addressed in the assessment of the efficacy of screening is "Would the screening decrease breast cancer mortality?"

The majority of the published reports found that screening for the breast cancer is beneficial for women. For instance, by analyzing the data from the Health Insurance Plan of Greater New York (HIP) collected during the 1960's.

Shapiro et al. (1982) concluded that by the end of 10 years after entry to screening study, breast cancer mortality was about 30% less in the study group than that in the control group. In another paper, Shapiro (1997) found that by the end of 18 years since entry to the study, the study group has about a 25% lower breast cancer mortality among women aged 40-49 and 50-59 (at time of entry) than that in the control group. Analyzing the same HIP data using a competing risks model, Aron and Prorok (1986) found that some but not all of the breast cancer cases detected early by screening realized a benefit in terms of reducing the breast cancer mortality. But screening appears not to affect the mortality rate from causes of death other than breast cancer.

Based on the NHS screening program conducted in England and Wales, Blanks et al. (2000) concluded that the total reduction of the breast cancer mortality for the study group was estimated as 21.3%.

Yet, there were researchers who reported that screening was not beneficial for women. Gotzsche et al. (2000) concluded that screening for breast cancer with mammography is unjustified. They reviewed each of the eight such screening trials that they could find in the public domain at the time, including the HIP study, and found that in six of the eight, the randomization process failed to create a study group similar to the control group, and that the remaining two trials, although adequately randomized were did not prove the effectiveness of screening.

Comparison between the study group and the control group is a very difficult task. Dividing individuals into a study group and a control group by randomization is conceptually simple but hard to carry out. Selection bias (not necessarily done knowingly) and noncompliance of the participants often pose serious problems for performing randomization. In the case of the HIP study, the question of why twice as many women in the study group as in the controlled group were excluded from participation due to their pre-existing breast cancer condition is still not resolved. If the screening trial were truly random, one would expect about the same number excluded from each group. The imbalance suggests that some sort of selection bias was there. It was also known in the HIP screening trial that a significant number of women refused to be screened at all while others who were willing but failed to show up for a scheduled screening. The noncompliance will not only affect randomization but also reduce the potential effectiveness of screening. These problems cast doubts on the comparability of the study group with the control group.

In addition to the randomization problems, the selection of the outcome variables (or endpoints), cohorts, and statistical methods for analysis also contributed to the conflicting research results. For example, some researchers measured a patient's survival time from the date of entry to the time of death, while others used the time of diagnosis as the initial time. Another major reason for different

findings is due to the selection of time period for comparing survival or mortality of the breast cancer patients.

Lastly, the heterogeneity of the patients makes the data analysis difficult. One could reduce the heterogeneity by limiting the investigation to a specific age-demographic cohort, but then sample size would become too small for analysis.

In this thesis we study the efficacy of screening using the HIP data with special attention to the non-compliance problem, the selection of cohorts for comparing the study group and the control group, and the selection of outcome variables.

The endpoint results collected by HIP are trivariate. It consists of death from breast cancer, death from other causes and survival of breast cancer of all those cancer cases detected during the entire period of the screening trial. If the endpoint results were bivariate with only the death from breast cancer and the survival of breast cancer, then the comparison using the breast cancer death proportions would be straight forward. In the trivariate situation, the question of "How do we use the information on death from other causes" needs to be carefully addressed.

In Shapiro et al. (1982), the breast cancer mortality was computed in two different ways, either (a) using only the breast cancer deaths and separately considering the mortality from other causes or (b) including the deaths from other causes into the category of deaths from breast cancer. On the contrary, we believe that the death from other causes should be considered as "censored"

survival times and hence to be included in the category of "survival of the breast cancer". Excluding or including the category of "death from other causes" as in approaches (a) and (b) would bias the calculation of the breast cancer mortality.

If screening is beneficial, one would expect a higher curing percentage of the breast cancer in the screening group than that of the control group. (We should note that curing means the cancer is in remission.) Therefore, by the end of the follow-up period, the screening group would have a lower proportion of breast cancer deaths and a higher proportion of deaths from other causes than that of the control group. Indeed this problem was investigated by Aron and Prorok (1986). However, our study differs from that of Aron and Prorok (1986) in the selection of cohort and the time period for comparison.

In this thesis we consider the trivariate outcome variables of the breast cancer patients detected in the study group and in the control group over the same period of 54 months (4.5 years) since the entry date of each participant. Four and half years is chosen to include the last screened-detected cancer case in the study group measured from her date of entry to the program. Our rationale is that after the cessation of screening (in 4.5 years), there were no more screening detected cancer cases in the study group. Therefore, the effect of screening should be investigated only during this period. Thus we do not include cancer cases detected in the ensuing follow-up period in the comparison. This is another difference between our investigation and that of the others. See, for example

Shapiro et al. (1982), Aron and Prorok (1986), Chu (1988) and Shapiro (1997). In the literature, cancer cases detected in various time periods, such as 5, 6, 7 and 10 years after the entry date have been used for comparing the study group and the control group. Some researchers chose a longer time period because of concern about false-negative screening results in the study group during the screening period and thus extended the time period for cancer detection to make allowance for "under detection" in the study group due to false negative result. However, lack of definitive conclusion about false negative result, this becomes a subjective choice.

We also have investigated five-year survival probability in this thesis. The five-year survival probability is the proportion of women alive five years after they were diagnosed with breast cancer.

A host of other issues arise in the analysis of the screening data. Among them we have addressed the problem of lead time.

The thesis is organized as follows. In Chapter 2, we give detailed descriptions of the HIP study. In Chapter 3, hypothesis testing is performed to test the equality of death proportions and survival proportions in the study group, control group and refused group. Next in Chapter 4, statistical tests concerning the difference in life expectancy among three groups were conducted. In Chapter 5, analysis of detection in early stages was conducted. Chapter 6 contains analysis of five-year survival probability. Analysis of the lead time is conducted in Chapter 7.

In Chapter 8, we propose a Markov chain for a future study and investigate what kind of data and sample size would be needed in future screening trials. The concluding remarks are given in Chapter 9.

Chapter 2

The HIP Randomized Controlled Trial and Selection of Cohorts

The Health Insurance Plan (HIP) study was initiated in December 1963 in the state of New York. The primary objective of the HIP study was to determine if periodic screening for breast cancer with mammography and clinical breast examination is beneficial for women.

In the beginning of the study, 62,000 women between ages 40 and 64 years with at least one year membership in HIP were randomly assigned to the study group and the control group. The randomization was performed in such a way that every other women in the data base was assigned to the control group which split the total number into two equal groups with 31,000 women in each group. However, 869 women in the study group and 435 women in the control group were excluded from the trial because those women were identified as having a prior breast cancer diagnosis. This left the HIP study with 30,131 women in the control group and 30,565 women in the study group. Each woman in the study

group was offered an initial screening for the breast cancer and three annual screening afterwards. However, 9,984 women assigned to the study group refused to participate in the screening. This reduced the sample size of the study group to 20,147 women. Those women who refused screening were also followed up and their data will be analyzed separately. The reason for conducting a separate analysis on the refused group is that it has been well established in the literature that women in the refused group have different characteristics. Thus the break down of the total number of participants is as follows:

Control group	30,565
Study group	20,147
Refused group	9,984

The follow-up of the entire control group, the study group and the refused group for breast cancer incidence and general mortality were conducted in the form of mail surveys 5, 10 and 15 years after after entry date for each woman. There were several women lost to follow-up, but fortunately the subset data used in this thesis does not have any woman lost to follow-up.

2.1 Screening and Cancer Detection

As mentioned above, women in the study group were offered an initial breast cancer screening examination and three subsequent annual breast cancer screening examinations.

Each examination consisted of a film mammography (cephalocaudal and lateral views of each breast), a clinical examination of the breast by a physician (usually a surgeon) and an interview for demographic and other background information. The mammography and the clinical examination were done separately/independently and later the findings were coordinated for reports to the women and their personal physician.

In addition to these four screening examinations, if the screening examination of a woman showed a questionable result, there could be a strong indication of cancer but not sufficient evidence to declare cancer, that particular woman would be recalled for early screening examination.

For women who were diagnosed with the breast cancer, the following data were collected with regard to when the breast cancer was diagnosed and whether it was attributed to screening.

1. Initial screening
2. Initial screening on early recall
3. First annual screening
4. First annual screening on early recall
5. Second annual screening
6. Second annual screening on early recall

7. Third annual screening

8. Third annual screening on early recall

9. Last exam was initial

10. Last exam was first annual

11. Last exam was second annual

12. Last exam was third annual

13. No histologic confirmation prior to death

Detection types from item 1 to 8 were attributed to the breast cancer screening examination, but not types 9 to 13. Types 9-13 mean that the breast cancer was detected between screenings and they are called interval detected cancer cases. Interval detection may occur due to a variety of reasons, including cancer was too small to be detected at the time of screening, false-negative result at one or more previous screenings, noncompliance by the participants, such as missing a scheduled screening or not following a recommendation for a biopsy/aspiration given after an examination, or simply dropping out the study. The issue of noncompliance will be discussed in Section 2.4.

Because the focus of this thesis is to access the efficacy of breast cancer screening, ideally we would like to use cases that were attributed to screening only. However, the issue of noncompliance make the task of separating cases that were

attributed to screening and cases that were not attributed to screening very difficult. Thus we decide to use all cancer cases.

Information from different sources was used to identify cause of death in all three groups. These sources included HIP records, hospital claim files, death records in several states, cancer registry for New York State, the National Death Index and the mail surveys. Cause of death was determined by reviewing death certificates and hospital and physicians' records. The reviewers were blinded to which group were the women in.

2.2 Selection of Cohorts for Comparison

The reason for conducting a screening trial is to assess whether periodical screening lead to detection of cancer at its early stages of growth, thereby increasing the chance of patient's survival and thus decreasing the breast cancer mortality.

By design of the HIP screening trial, women in the study group were offered four screening examinations for the breast cancer. This took place in the first 4.5 years of the screening trial. After this period no woman in the study group were offered any additional screening. Thus to compare the study group and the control group, we use only the cancer cases detected during this 4.5 year period in all groups as our cohorts.

The HIP study recorded all breast cancer cases during the length of study which include the screening period and the follow-up period. The cancer cases detected during the follow-up period will be studied separately.

We should note that the determination of the actual time of the screening period of 4.5 years is not without complications. There are several reasons why we choose 4.5 years.

The entry date to the HIP study for each woman were varied. For women in the study group, the entry date can be different from the initial screening examination date. Moreover, although the screening period for each woman in the study group should be three years with about one year difference between annual screening examinations, the HIP data revealed a rather different picture. The annual schedule was not always followed. In a number of cases the time between scheduled annual screening examination greatly surpassed one year. In fact the last screen-detected breast cancer occurred in the 50^{th} month since her entry date. This was 4.17 years after the entry, not the 3 years as scheduled. Table 2.1 presents the data of the time between entry to initial screening examination and also between screening examinations.

As can be seen in Table 2.1, it took 66 months or 5.5 years from the entry date for women in study group to complete all screening examinations. However, we believe that using 66 months to determine the screening period is not reasonable because the screening period would then be stretched to an unreasonable length.

Table 2.1: Time in Months from Entry to Initial Screening and between Screening Examinations

	Average	Std Dev	Min	Max
Entry to initial screening	0.78	1.502	0	18
Initial to 1^{st} annual screening	13.50	1.756	7	23
1^{st} annual to 2^{nd} annual screening	13.42	1.615	1	23
2^{nd} annual to 3^{rd} annual screening	13.69	1.881	8	31
Entry to 3^{rd} annual screening	40.99	3.386	32	66
Initial to 3^{rd} annual screening	40.36	2.966	25	62

Std Dev = standard deviation, Min = minimum, Max = maximum

Table 2.1 shows that it took 18 months for all women who entered the study group to complete the initial screening. In theory, it should take 36 months for each woman to finish all three annual screening examinations. Therefore we believe 54 months (18 + 36 months) or 4.5 years is a reasonable screening period for the study group. Fortunately, the last screening detected case was diagnosed in the 50^{th} month after entry date, which is within the 54 months that we have chosen. Thus, a 54 month screening period will be used in the comparison of the study group, the control group and the refused group with respect to cancer detection, mortality and survival. This establishes the cohorts for assessing the efficacy of screening. Table 2.2 compares the breast cancer cases diagnosed during the entire HIP study and during our screening period.

Table 2.2: Comparison of Number of Breast Cancer Cases Diagnosed during HIP Study and Screening Period

Group	Total participants	Number of cancer cases diagnosed during			
		HIP study		screening period	
		No of cases	Percentage	No of cases	Percentage
Control	30,565	945	3.092%	268	0.877%
Study	20,147	681	3.380%	211	1.047%
Refused	9,984	280	2.804%	71	0.711%

The selection of time period for comparison of the study group and the control group is crucial to the analysis. Other researchers who used the HIP data have chosen different time periods in their analysis. Chu et al (1988) used 6 years since entry as opposed to our 4.5 years. Their reason was that 6 years was the earliest time at which the number of breast cancer cases diagnosed in the control group and the study group were equal. Shapiro (1989) used all cancer cases diagnosed in 5 and/or 7 years after the entry date as the research subjects. Other periods such as 10 years were also used in the literature. See, for example Shapiro et al (1982). Inclusion of cancer cases detected during the follow-up period (not obtained by screening) would water down the effect of screening and hence bias the assessment of the efficacy of screening.

2.3 Detection Stage

In the HIP study, the detection stage of each woman was recorded. The criteria used for classification is as follows:

Stage I No microscopic evidence of axillary lymph node metastasis; no skin involvement; no pectoral muscle or chest wall attachment; no distant metastasis.

Stage II Positive microscopic evidence of axillary lymph node metastasis; no skin involvement; no pectoral muscle or chest wall attachment; no distant metastasis.

Stage III Axillary lymph nodes may be microscopically negative, positive or unknown; evidence of skin involvement - erythema, infiltration, ulceration, peau d'orage, edema; or pectoral or chest wall attachment; or clinically palpable supraclavicular lymph node(s) with positive microscopic evidence of metastasis; or axillary lymph nodes fixed to one another or to other structures; no distant metastasis.

Stage IV Positive evidence of distant metastasis, radiographically or histologically confirmed.

In addition to these 4 stages, there are two other classifications used in the HIP study. One is "Clinical Stage I or II", the other is "unknown".

Clinical stage I or II is an invented stage, there were no axillary lymph nodes removed. The only way to distinguish between stage I and stage II is by looking at the lymph nodes. Since there were no axillary lymph nodes removed, there was no way of knowing for sure whether it was stage I or stage II. However, there was no demonstrable evidence of stage III or stage IV cancer. Thus it could not be stage III or stage IV.

Women whose detection stage were unknown are categorized as "unknown". There are some examples of such cases in Section 2.4.

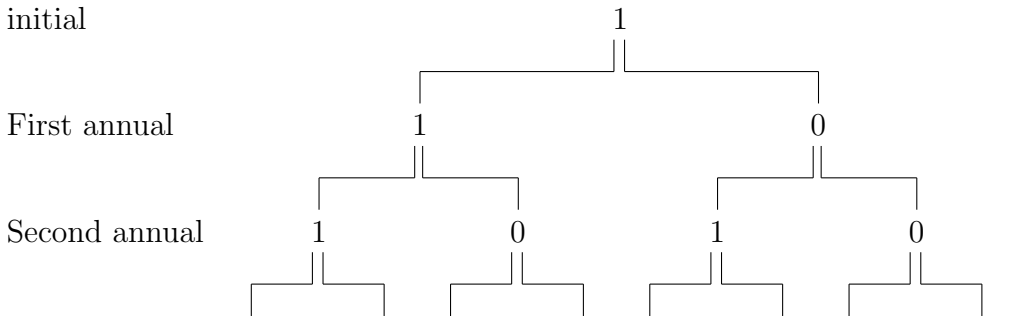
There are only a few clinical stage I or II cases, thus for analysis purpose clinical stage I or II detections were combined with stage I detection.

2.4 Compliance of Women in the Study Group

In this section, we examine data on the compliance of women in the study group. The screening pattern of all women in the study group is given in Figure 2.1.

Note that women in the study group who were diagnosed with breast cancer on one of the screening examinations or between examinations would not attend the ensuing screening examination. For instance, a woman who was diagnosed with the breast cancer on the initial screening would have no need to come to any of the subsequent annual screenings. The proportion of non compliance were affected by this. However, it should be noted that the proportion of all cancer cases in the study group is 3.380%, which is relatively small.

Figure 2.1: Compliance of All Women in the Study Group



Pattern	1111	1110	1101	1100	1011	1010	1001	1000
No of women	11963	1715	793	1461	663	422	484	2646
Total	20147							
Percentage	59.38%	8.51%	3.94%	7.25%	3.29%	2.09%	2.40%	13.13%

It is very interesting to note that four cases in the study group have no histologic confirmation prior to death, that is the cancer stage at detection is unknown, even though the cause of death was breast cancer. Further investigation revealed that one woman only went to the initial screening, one woman went to the initial and first annual screenings, one woman went to the initial, first and second annual screenings and one woman went to all four screenings. These women were recommended to have biopsy or aspiration but none of them did it. Breast cancer was listed as their as cause of death.

From Figure 2.1, one sees that aside from a handful number of women who were diagnosed with breast cancer on screening or on early recall, only 59.38% (11,963 out of 20,147) women in the study group went to all four screening ex-

aminations. Initially, there were 30,131 women in the supposed study group. If we take the proportion of women who attended all four screening examinations with respect to these 30,131 women, only about 39.70% of women in the supposed study group complied with screening protocol. Future studies on efficacy of breast cancer screening should take this drastic reduction in sample size in the study group into consideration. In the screening design, more women should be assigned to the the study group in order to maintain a reasonable sample size for the study group for comparison with the control group.

2.5 Summary of the Data of the Control, Study and Refused Groups

Control Group Women in this group did not receive screening examinations. There are 30,565 women participants in this group. Among them, 268 women were diagnosed with breast cancer during the next 4.5 years after entry which corresponded to the screening period for the study group. In contrast 945 were diagnosed with cancer during the entire 18 years of the HIP study. Among these 268 cancer cases, by the end of the HIP study 97 survived, 147 died from breast cancer and 24 died from other causes causes. The average entry age of the control group is 50.90 years with a standard deviation of 6.577 years, while the average entry age of cancer

cases diagnosed during screening period in the control group is 52.31 years with a standard deviation of 6.587 years.

Study Group Women in this group received screening examinations. There are 20,147 women in the study group. Among them, 211 women were diagnosed with breast cancer during the screening period. In contrast 681 women of this group were diagnosed during the entire HIP study. Out of 211 cancer cases in this group, by the end of the HIP study 93 survived, 84 died from breast cancer and 34 died from other causes. The average entry age of the whole study group is 50.74 years with a standard deviation of 6.524 years, while the average of entry age of cancer cases diagnosed during screening period in the study group is 52.47 years with standard deviation of 6.501 years.

Refused Group Women in this group were offered but refused screening examinations. There are 9,984 women in this group. During the screening period 71 women were diagnosed with the breast cancer. In contrast, during the whole HIP study, 280 women were diagnosed with cancer. Out of 71 cancer cases, by the end of the HIP study 23 survived, 33 died from breast cancer and 15 died from other causes. The average entry age of the whole the refused group is 51.11 years with a standard deviation of 6.608 years, while the average entry age of cancer cases detected during screening period in the refused group is 51.55 years with a standard deviation of 6.129 years.

Table 2.3: Number Cancer Cases during Screening Period by Status at the End of HIP Study

Group	Total cases	Survival		Breast cancer death		Other causes death	
		N	Percentage	N	Percentage	N	Percentage
Control	268	97	36.19%	147	54.85%	24	8.96%
Study	211	93	44.08%	84	39.81%	34	16.11%
Refused	71	23	32.39%	33	46.48%	15	21.13%

N = no of cases

Table 2.3 and Table 2.4 present brief summaries of these three groups.

It is important to note that HIP has no record of the remission or recurrence of those cancer cases. Thus for women who died from breast cancer, we have no knowledge whether the breast cancer is a first-case breast cancer or a re-occurring cancer. Also, for women who survived breast cancer, we do not know if the cancer was in remission or has recurred.

As mentioned above, the entry age of the women participating in this study was between 40-64 years with an average of 50.88 years and a standard deviation of 6.565 years. The length of the HIP study itself is around 18 years. By the end of the HIP study, the age of the women participants was around 58-82 years. Thus some participants are quite advanced in age and this can affect their health situation as a whole.

Table 2.4: Average and Standard Deviation of Entry Age in Year

Group	All Women			Cancer cases during screening period		
	N	Average	Std Dev	N	Average	Std Dev
Control	30,565	50.90	6.577	268	52.31	6.587
Study	20,147	50.74	6.524	211	52.47	6.501
Refused	9,984	51.11	6.608	71	51.55	6.129

Std Dev = standard deviation

We have compared the entry age and also compared the proportions of the breast cancer cases diagnosed during screening period among groups.

We use ANOVA to test if the average entry ages are significantly different among groups. If so, Tukey's confidence intervals are produced to investigate in which pair of groups the different exist.

The first ANOVA is conducted using the entire HIP data. It shows that differences of the average entry ages among groups is significant at 5% level (the F statistic is 10.54 with 2 degrees of freedom which gives a p-value of < 0.0001). Using Tukey's simultaneous confidence intervals, we find that the average difference of entry age between the control group and the study group is 0.160 year with 95% confidence interval of (0.020, 0.300), the average difference of entry age between the control group and the refused group is -0.204 year with 95% confidence interval of (-0.381, -0.027) and the average difference between the

study group and the refused group is -0.364 years with 95% confidence interval of (-0.552, -0.176).

In contrast, the second ANOVA applied to entry age of the breast cancer cases diagnosed during screening period shows no significant entry age differences among three groups (the F statistic is 0.54 with 2 degrees of freedom which gives a p-value of 0.5825).

The proportions of breast cancer diagnosed during the screening period in the control group, the study group and the refused group are 0.877%, 1.047% and 0.711% respectively. The pairwise comparisons show that:

(1) The difference of proportions between the control group and the study group is -0.0017 with a standard error of 0.0009 and the an 95% asymptotic confidence interval of (-0.0035, 0.0000). Since the confidence interval contains zero, the proportion difference is not significant at 5% level.

(2) The difference of proportions between the control group and the refused group is 0.0017 with a standard error of 0.0010 and an 95% asymptotic confidence interval of (-0.0003, 0.0036). The difference is not significant.

(3) The difference of proportions between the study group and the refused group is 0.0034 with a standard error of 0.0011 and a 95% asymptotic confidence interval of (0.0012, 0.0055). Hence the proportion difference is significant.

Even though the proportion difference between the control group and the study group is not significant, there were indeed more breast cancer cases detected in the study group compared to the control group. This possibly indicates that annual screening causes a shift in the detection time.

Throughout the thesis, it is to be understood that the study sample is limited to those women whose breast cancer was detected during this screening period, unless noted otherwise. Also, all tests are conducted with a 5% level of significance, unless noted otherwise.

Chapter 3

Comparison of Proportions of Cancer Mortality, Survival and Deaths from Other Causes

As discussed in Chapter 1, in any group each cancer patient had one of three possible outcomes by the end of the HIP study. A patient can be either alive, or die from breast cancer, or die from other causes. Although there were several participants lost to follow-up in the HIP project, fortunately there are no women lost to follow-up for the particular data set used in this thesis.

In this chapter, we compare the proportions of women in each of the three possible outcomes among the control group, the study group and the refused group. If screening examination does lead to early detection of the breast cancer which in turn lead to better survival probability, we would then see a larger proportion of breast cancer survivors in the study group as compare to the control group and probably the refused group. We would also expect to see a smaller proportion of breast cancer deaths in the study group than in the control group and probably the refused group. As for deaths due to other causes, it is likely

that we would see a larger proportion of it in the study group than in the control group because those cancer survivors in the study group would die from other causes.

The proportions of the three possible outcomes for all three groups in the HIP data were given in Table 2.3.

3.1 Test of Equality of Proportions

We test the hypothesis of equality of proportions based on a multinomial distribution of three possible outcomes at the end of the HIP study. The test will then be followed by pairwise comparison tests of equality of proportions. The result will give us a sense as to which pairs are different.

We first compare all three groups. Let p_{A-C}, p_{A-S} and p_{A-R} be the respective true proportions of women who survived breast cancer in the control group, the study group and the refused group by the end of the follow-up. Similarly, let p_{B-C}, p_{O-C}, p_{B-S}, p_{O-S}, p_{B-R}, p_{O-R} denote the respective proportions of women who died from breast cancer and from other causes in the control group, the study group and the refused group. Then the hypothesis to be tested is

$$H_0 : p_{A-C} = p_{A-S} = p_{A-R},\ p_{B-C} = p_{B-S} = p_{B-R},\ p_{O-C} = p_{O-S} = p_{O-R}$$

versus the alternative

$$H_A : H_0 \textit{ is not true.}$$

The chi-square statistic yields 16.6877 with 4 degrees of freedom which gives a p-value of 0.0022. Thus we reject the null hypothesis of equal proportions at the 5% level.

However, this test does not give us a sense as to which pairs are different. Thus we conduct a pairwise comparison between each pair of two groups. Since each group is used twice in the pairwise comparison, the resulting three tests involved the same data sets and hence are not statistically independent. By invoking Bonferroni's inequality, if we set the level of significance of each separate pairwise test to be $0.05/3 = 0.0167$, we can claim that the level of significance of the test of equality of all three pairs is no larger than 0.05.

We first compare the control group and the study group. the hypothesis to be tested is $H_0 : p_{A-C} = p_{A-S},\ p_{B-C} = p_{B-S},\ p_{O-C} = p_{O-S}$ versus the alternative H_A : H_0 *is not true*. The chi-square statistic yields 12.383 with 2 degrees of freedom which gives a p-value of 0.002. We reject the null hypothesis of equal proportion at level 0.0167.

Similar tests were also conducted to compare the control group and the refused group, and also the study group and the refused group.

The test shows that the control group and the refused group have significantly different proportions. The chi-square statistic is 8.198 with 2 degrees of freedom which gives p-value of 0.0166.

In comparison of the study group and the refused group, the chi-square statistic is 3.100 with 2 degrees of freedom which gives a p-value of 0.2122. Thus the null hypothesis of equal proportions is not rejected at level 0.0167.

Based on these results, we can say that the refused group is "closer" to the study group than to the control group. This conclusion is surprising since this seems counter intuitive. Intuitively, since the women in the refused group did not have screening, we expected the result to be similar to that of the control group.

Further tests of proportions were performed to investigate in which direction the proportions differ.

3.2 Separate Comparisons of Single Outcome Variables

We will compare each of the three outcome variables separately the in two groups (pairwise) at a time. We will compare the control group and the study group, and compare the control group and the refused group. Since the test in the last section did not reject the equality of two trinomial distributions of proportions of the study and refused groups, we will not conduct a pairwise comparison of the study group and the refused group.

The deaths from other causes complicate our comparisons of the breast cancer survival proportion and breast cancer death proportion. The HIP data does not

contain any information about the curing (cancer in remission) and reoccurring of the breast cancer and thus it is not possible to ascertain the disease status (with respect to the breast cancer) of those women who died from other causes. In the absence of such information, we consider two approaches. In the first approach, we treat each outcome variable separately. In the second approach, we treat the data of death from other causes as censored survival data because they were diagnosed with cancer and were alive until their death due to other causes. Thus their lifetimes are censored observations with respect to the breast cancer survival.

3.2.1 Comparison of Breast Cancer Mortality in Different Groups

Referring to Table 2.3, the proportions of death from breast cancer among women diagnosed with the breast cancer in the control group, the study group and the refused group are 0.5485 (147/268), 0.3981 (84/211) and 0.4648 (33/71) respectively.

The binomial test shows a significant difference in the proportion of breast cancer death in the control group and the study group. The proportion of breast cancer death is 0.1504 lower in the study group. The standard error of the difference is 0.0454 with an asymptotic 95% confidence interval of (0.0615, 0.2394).

The Mantel-Haenszel relative risk for these two groups is 1.3778 with 95% confidence interval of (1.1300, 1.6800). This means that women diagnosed with the breast cancer during screening period in the control group are 38% more likely to die from breast cancer than those women in the study group by the end of the follow-up.

However, the test shows no significant difference in the breast cancer death proportion between the control group and the refused group at 5% level.

3.2.2 Comparison of Proportion of Death from Other Causes

Referring to Table 2.3, the proportions of death from other causes in the control group, the study group and the refused group are 0.0896 (24/268), 0.1611 (34/211) and 0.2113 (15/71) respectively.

The difference in proportions between the control group and the study group is significant. The proportion of deaths from other causes is 0.0716 higher in the study group. The standard error of the difference is 0.0307 with an asymptotic 95% confidence interval of (0.0113, 0.1318).

The difference in proportions of deaths from other causes is significant between the control group and the refused group. The proportion of deaths due to other causes is 0.1217 higher in the refused group. The standard error of the difference is 0.0515 with an asymptotic 95% confidence interval of (0.0208, 0.2226).

It is important to note that there are 7.16% more women in the study group who died from other causes compared to the control group. As mentioned previously, this can be interpreted as saying that early detection of breast cancer by screening in the study group results in higher probability of recovery (or cancer in remission) from the breast cancer but then those women eventually died from other causes.

3.2.3 Comparison of Survival Proportions

From Table 2.3, the proportions of women who survived the breast cancer in the control group, the study group and the refused group are 0.3619 (97/268), 0.4408 (93/211) and 0.3239 (23/71) respectively.

Even though the survival proportion in the study group is 0.0788 higher compare to that in the the control group, the test shows that the result is not significant at 5% level. The standard error is 0.0451 with an asymptotic 95% confidence interval of (-0.0095, 0.1671). Although the result is not significant, the lower bound of the confidence interval, which is -0.0095, is quite close to zero.

Also, there is no significant difference between the control group and the refused group proportion at the 5% level.

3.2.4 Treatment of the Death Due to Other Causes

In this subsection, we consider two different ways of treating the death due to other causes. Our first approach is to completely ignore the category of death from other causes in the calculation of the breast cancer mortality. Our second approach is to treat the times to death due to other causes as censored survival times. This is because an individual who died from other causes was alive as a breast cancer patient until her time of death, and her survival time was then censored at the time of her death. Either approach will give us two mutually exclusive outcome events. A women would then either die from breast cancer or survive breast cancer with a censored survival time. Thus comparing survival proportion is equivalent to comparing the breast cancer death proportion. We shall compare the survival proportion. Note that either approach would result in a modification of the original data.

The first approach will reduce sample sizes in the three groups, so they need to be recalculated. Referring to Table 3.1, in the control group 244 women were diagnosed with breast cancer during screening period. By the end of the HIP study, there were 97 survivors and 147 women who died from breast cancer. While in the study group, 177 women were diagnosed with breast cancer during the screening period of whom 93 survived and 84 died from the breast cancer. Lastly, in the refused group, there were 23 women who survived and 33 who died from breast cancer, making a total of 56 women.

Table 3.1: Number Cancer Cases during Screening Period by Status at the End of the Follow-up

Group	Total no of cases	No of survival	No of breast cancer deaths
Control	244	97	147
Study	177	93	84
Refused	56	23	33

The survival proportions in the control group, the study group and the refused group then become 0.3975, 0.5254 and 0.4107 respectively.

The control group and the study group showed a significant result at 5% level. The survival proportion in the study group is 0.1279 (standard error is 0.0489) higher than in the control group. An asymptotic 95% confidence interval for the difference of survival proportions is (0.0321, 0.2237). Furthermore, the Mantel-Haenszel relative risk for these two groups is 1.2695 with the 95% confidence interval of (1.0545, 1.5282). In other words, it can be said that women diagnosed with breast cancer during screening period in the study group are 27% more likely to survive breast cancer than those women in the control group.

For the study group and the refused group comparison, and also the control group and the refused group comparison, the tests show no significant differences at 5% level.

Table 3.2: Number Cancer Cases during Screening Period by Status at the End of the Follow-up

Group	Total no of cases	No of survival	No of breast cancer deaths
Control	268	121	147
Study	211	127	84
Refused	71	38	33

The second approach for comparing proportions of women who survived breast cancer is conducted by combining death times from other causes with the survival times. Table 3.2 gives the recalculated number of cases. Under this approach, in the control group there were 121 women survived breast cancer (censored survival times) and 147 who died from breast cancer which gives a total of 268 women diagnosed with the breast cancer. In the study group there are 211 women diagnosed with breast cancer of whom 127 survived and 84 died. Lastly, in the refused group, there were a total of 71 women diagnosed with breast cancer of whom 38 survived and 33 died.

The survival proportions in the control group, the study group and the refused group are 0.4515, 0.6019 and 0.5352 respectively.

For the control group and the study group comparison, the test shows a significant difference at 5% level. The survival proportion in the study group is 0.1504 (with standard error 0.0454) higher than in the control group. A 95% confidence

interval for this difference is (0.0615, 0.2394). The Mantel-Haenszel relative risk for these two groups is 1.3778 with 95% confidence interval of (1.1300, 1.6800). Thus we can also conclude that women in the study group are 38% more likely to survive cancer than those in the control group.

The test showed no significance difference between the study group and the refused group, and also no significance difference between the control group and the refused group.

Chapter 4

Comparison of Total Years Lived and Kaplan-Meier Estimate of Survival Function

In this chapter, we shall assess the effect of screening in terms of the total years lived and the Kaplan-Meier estimate of the survival function.

4.1 Comparison of Total Years Lived

For women who survived breast cancer by the end of the follow-up, the total years lived is defined as the length of time from the diagnosis of breast cancer to the end of the follow-up. For women who died either from breast cancer or from other causes before the end of the HIP study, the total years lived is defined as the length of time from the diagnosis of breast cancer to the time of death.

Since the age range (from 40 to 64 years of old) of women in the HIP study is fairly large, the overall (or natural) death rate increases significantly with the advancing age. To account for the age differences, another way of defining the

total years lived is used by some researchers. It is measured from the time of birth to either the time of death or the end of the follow-up if the woman is still alive.

In either case, if screening is beneficial, we would expect that the average of total years lived in the study group to be longer than that of the control group and probably the refused group. However one should keep in mind that the total-years-lived is measured up to the end of the follow up time. It does not represent the total lifetime of a patient

Comparison of the total years lived is complicated by the fact that a woman may die from other causes. Because HIP has no record of these women with respect to their breast cancer history and survival experience, we shall treat this type of death times as censored survival times.

Comparisons of the averages of total years lived in the three groups are carried out by using ANOVA. If ANOVA shows a significant result, then Tukey's construction of confidence intervals of all possible pairwise average differences between groups will be calculated, namely the difference between the study and control groups, the study and refused groups and the control and refused groups. These confidence intervals will be referred to as simultaneous confidence intervals. They are calculated to examine in which pairs the difference exist.

A five percent level of significance will be used throughout and hence it will not always be mentioned.

Table 4.1: Average and Standard Deviation of Total Years Lived after Detection for All Cancer Cases

	All cases		
Group	Number of cases	Average	Std Dev
Control	268	9.00	6.760
Study	211	11.11	6.042
Refused	71	9.03	6.755

Std Dev = standard deviation

4.1.1 Comparison of Total Years Lived after Cancer Detection

An ANOVA of the data presented in Table 4.1 shows that there is a significant difference among the three groups in the averages of the total years lived after detection (to be called averages for short in this section whenever no ambiguity arises). The F statistic is 6.86 with 2 degrees of freedom which gives a p-value of 0.0011. The Tukey pairwise test showed that the average difference between the control group and the study group is significant. The average in the study group is 2.1158 years longer than that in the control group. The Tukey 95% confidence interval for the true average difference is (0.7114, 3.5203). On the other hand, the pairwise test shows no significant difference in averages between either the control group and the refused group or the study group and the refused group.

Note that at the end of HIP study there are three possible outcomes for women diagnosed with the breast cancer: survival, death from breast cancer or death from other causes. Since there is no information recorded on the cancer development of those women who died from other causes, we are again faced with the problem of how to properly utilize the death times of these women. We decided to treat them as a separate case. The following refinement of ANOVA is conducted separately for deaths due to breast cancer, survivals and combination of deaths due to breast cancer and survivals outcomes. Table 4.2, Table 4.3 and Table 4.4 contain the averages and standard deviations for various combinations of group and outcome. Table 4.4 excludes deaths from other causes. The ANOVA was not conducted for deaths from other causes, because deaths from other causes give us no information on breast cancer survival experience. Even though no analysis was conducted for deaths due to other causes, the averages and standard deviations were given in Table 4.5.

Death from Breast Cancer The data are presented in Table 4.2. The ANOVA for death from breast cancer shows significant differences among the three groups at 5% level. The F statistic is 4.19 with 2 degrees of freedom which give a p-value of 0.0162. The Tukey pairwise comparison shows that a significant difference in the averages of the total years lived between the study group and the control group. The average of the study group is 1.4087 years longer than that of the control group with a 95% confidence interval

Table 4.2: Average and Standard Deviation of Total Years Lived after Detection for Women Who Died from Breast Cancer

Group	Breast cancer death		
	Number of cases	Average	Std Dev
Control	147	4.20	3.977
Study	84	5.61	4.173
Refused	33	3.68	3.988

Std Dev = standard deviation

Table 4.3: Average and Standard Deviation of Total Years Lived after Detection for Women Who Survived

Group	Survival		
	Number of cases	Average	Std Dev
Control	97	16.63	1.584
Study	93	16.62	1.449
Refused	23	16.63	1.508

Std Dev = standard deviation

Table 4.4: Average and Standard Deviation of Total Years Lived after Detection for Breast Cancer Death and Survival Combined

Group	Death from Breast Cancer and Survival		
	Number of cases	Average	Std Dev
Control	244	9.14	6.903
Study	177	11.40	6.303
Refused	56	9.00	7.175

Std Dev = standard deviation

Table 4.5: Average and Standard Deviation of Total Years Lived after Detection for Women Who Died from Other Causes

Group	Other Causes Death		
	Number of cases	Average	Std Dev
Control	24	7.49	4.941
Study	34	9.62	4.210
Refused	15	9.14	5.085

Std Dev = standard deviation

of the true average difference (0.1057, 2.7117). However, Tukey's method shows no significant difference in averages in all other pairwise comparisons of two groups.

Survival The data are presented in Table 4.3. ANOVA shows no significant differences in the averages of total years lived among the survivals in the three groups. The F statistic is 0.00 with 2 degrees of freedom which give a p-value of 0.9991. We expect this result because the women are all censored by end of follow-up.

Combined Data of Deaths from Breast Cancer and Survivals The data are presented in Table 4.4. The ANOVA for the two combined outcomes shows a significant difference in averages of total years lived since detection among three groups. The F statistic from the ANOVA is 6.42 with 2 degrees of freedom, which gives a p-value of 0.0018. The results of the Tukey pairwise comparison of the averages of total years lived are as follows. The average difference between the control group and the study group is significant at 5% level. The average of the study group is 2.2534 years longer than that of the control group with a 95% confidence interval of the true average difference (0.6935, 3.8132). The tests showed no significant differences in the averages of total years lived between the control and the refused group, and also between the study and the refused group.

Table 4.6: Average and Standard Deviation of Total Years Lived after Birth for All Cancer Cases

Group	All cancer cases		
	Number of cases	Average	Std Dev
Control	268	64.03	9.287
Study	211	66.02	9.036
Refused	71	63.38	8.077

Std Dev = standard deviation

4.1.2 Comparison of Total Years Lived Measured from the Time of Birth

The summary statistics of total years lived since the time of birth is given in Table 4.6. Data analyses similar to those in the previous section for the total years lived since cancer detection were performed.

The ANOVA for the averages of total years lived since birth (to be called averages in this section when no ambiguity arises) shows a significant difference among three groups at 5% level. The F statistic is 3.71 with 2 degrees of freedom which gives a p-value of 0.0250. The Tukey pairwise comparison shows that there is a significant difference in averages between the study group and the control group. The average of the study group is 1.9821 years longer than that of the control group. The 95% confidence interval for the true average difference is

Table 4.7: Average and Standard Deviation of Total Years Lived after Birth for Women Who Died from Breast Cancer

Group	Breast cancer death		
	Number of cases	Average	Std Dev
Control	147	58.77	7.354
Study	84	60.09	8.366
Refused	28	58.65	7.018

Std Dev = standard deviation

(0.0259, 3.9384). On the contrary, no significant differences were detected in other pairwise comparisons.

Table 4.7, Table 4.8, Table 4.9 and Table 4.10, contain the averages and standard deviations for different outcomes at the conclusion of the HIP study. As in Subsection 4.1.1, the ANOVA and Tukey pairwise comparisons of total years lived since birth are conducted for deaths due to breast cancer, survivals and combination of deaths due to breast cancer and survivals.

Death from Breast Cancer The ANOVA of the data in Table 4.7 shows no significant differences in averages of the total years lived since birth in the three groups. The F statistic is 0.89 with 2 degrees of freedom which gives a p-value of 0.4129.

Table 4.8: Average and Standard Deviation of Total Years Lived after Birth for Survival

Group	Survival		
	Number of cases	Average	Std Dev
Control	97	71.64	6.357
Study	93	71.13	6.377
Refused	23	69.25	5.363

Std Dev = standard deviation

Table 4.9: Average and Standard Deviation of Total Years Lived after Birth for Women Who Either Died from or Survived Breast Cancer

Group	Breast Cancer Death and Survival		
	Number of cases	Average	Std Dev
Control	244	63.89	9.398
Study	177	65.89	9.210
Refused	56	63.00	8.239

Std Dev = standard deviation

Table 4.10: Average and Standard Deviation of Total Years Lived after Birth for Women Who Died from Other Causes

Group	Other causes death		
	Number of cases	Average	Std Dev
Control	24	65.52	8.100
Study	34	66.67	8.167
Refused	15	64.79	7.537

Std Dev = standard deviation

Survival Referring to Table 4.8, The ANOVA shows no significant differences in averages among groups. The F statistic is 1.35 with 2 degrees of freedom which give a p-value of 0.2607.

Combined Data of Death from Breast Cancer and Survivals The data are presented in Table 4.9. The ANOVA shows that there is a significant difference in the averages of the total years lived since birth among three groups. The F statistic is 3.30 with 2 degrees of freedom which gives a p-value of 0.0376. However, there are no significant results from the Tukey pairwise average comparison between any two groups.

We have some very interesting facts here. There are significantly fewer breast cancer deaths and more breast cancer survivors in the study group than that of the control group. But the number of deaths :from other causes is significantly

higher in the study group than that in the control group. A plausible explanation is that women in the study group survived breast cancer but then died from other causes and hence had a higher mortality from other causes. However, neither the average of total years lived after detection nor the average of total years lived since birth of cancer survivors in the study group is significantly different from those in the control group. As for women who died from breast cancer, the averages of total years lived after birth are about the same in both the study and the control groups. But the study group has a significantly longer average of total years lived after detection than that of the control group.

A plausible explanation is that those who survived breast cancer in the control group were diagnosed at an early stage as was expected to happen in the study group, thus the total years lived after detection and the total years lived after birth are not significantly different. It would be interesting to investigate what happens when the analysis is conducted by the stage of the breast cancer at detection time.

4.2 Kaplan-Meier Estimator of Survival Function

4.2.1 Estimator of the Survival Function of the Breast Cancer Patients

A cancer patient may either die from breast cancer or from other causes or be alive at the end of the follow-up period. We are interested in the time X, which measures the time of a cancer patient from diagnosis to death from breast cancer. If the cancer patient dies from other causes or is alive at the end of the follow-up, let Y denote either the time from her diagnosis to death of from other causes or to the end of the follow-up if she is still alive. Here, we treat Y as a right censoring variable for X. The right censored data are $\min(X, Y)$ and the indicator $I[X \leq Y]$. The Kaplan-Meier estimate of the survival function $S(t) = P[X > t]$ of X for each group constructed from the censored data is given in Figure 4.1. Figure 4.2 shows the censoring time for each group.

Confidence bands for the Kaplan-Meier estimate of the survival distribution of X are produced using the Hall-Wellner method and are given in Figure 4.3, Figure 4.4 and Figure 4.5.

Figure 4.1 shows that the study group always has a better survival probability than that of the control group and the refused group. We conduct a generalized

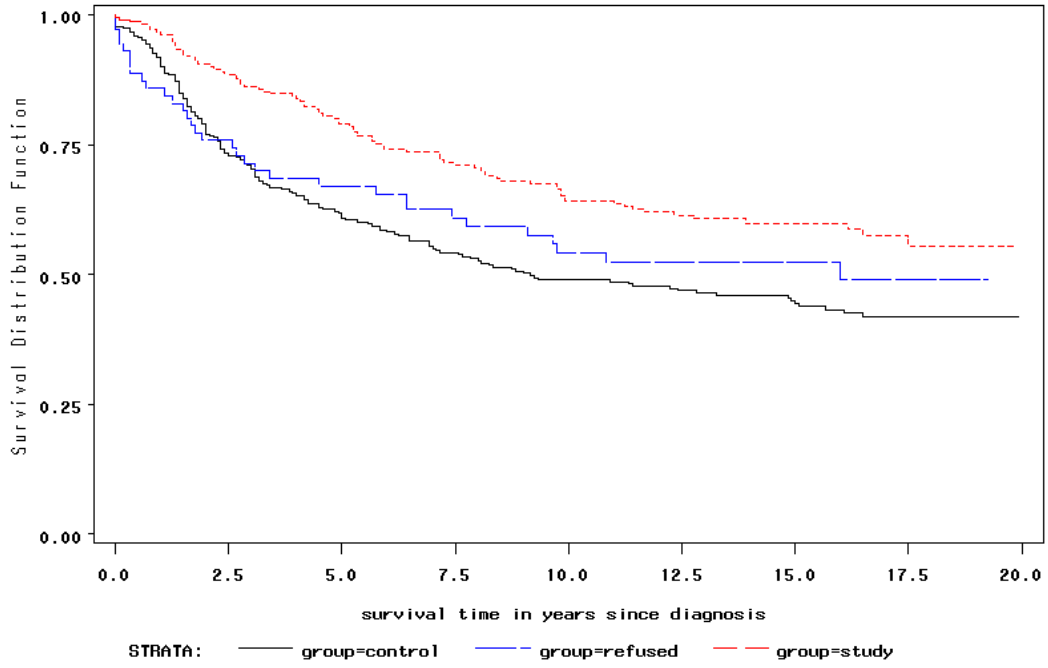


Figure 4.1: Kaplan-Meier Estimates of Survival Functions

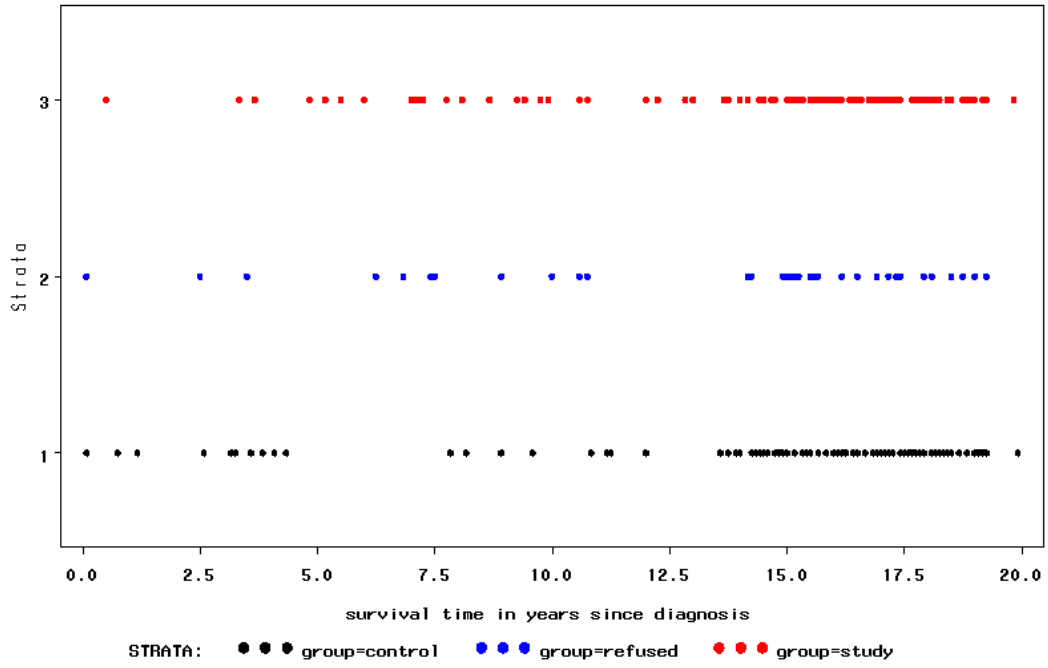


Figure 4.2: Censoring Pattern

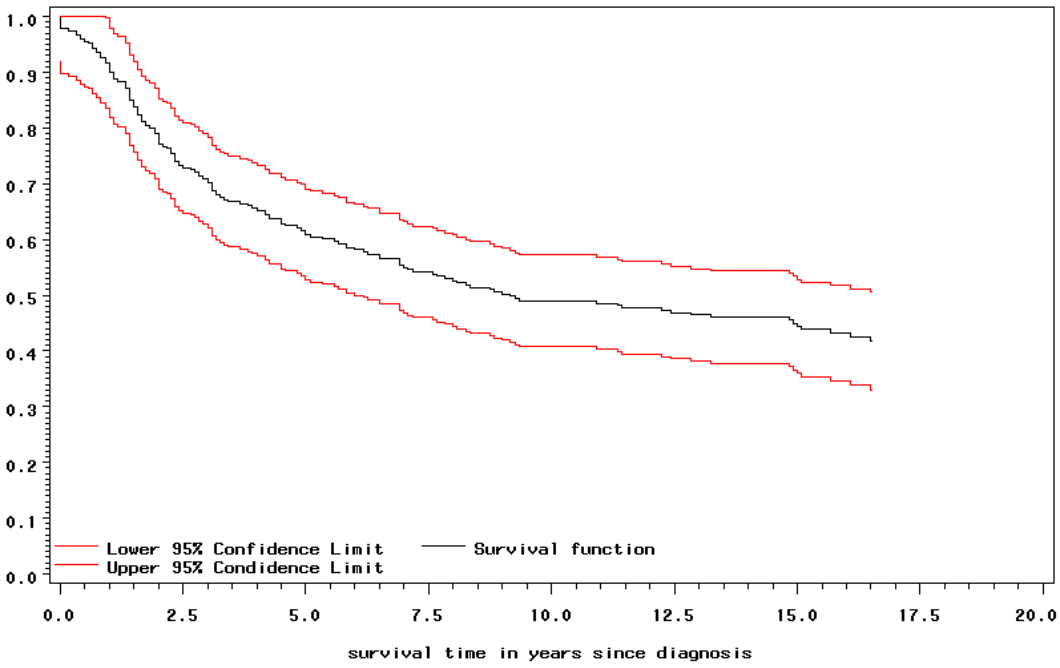


Figure 4.3: Hall-Wellner Confidence Bands for the Control Group

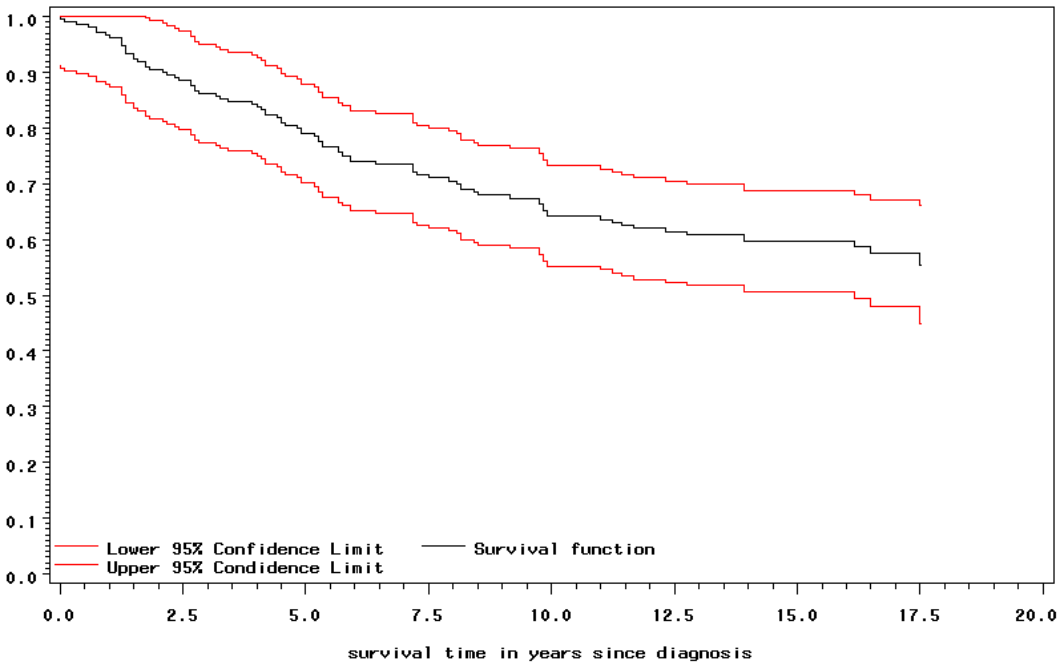


Figure 4.4: Hall-Wellner Confidence Bands for the Study Group

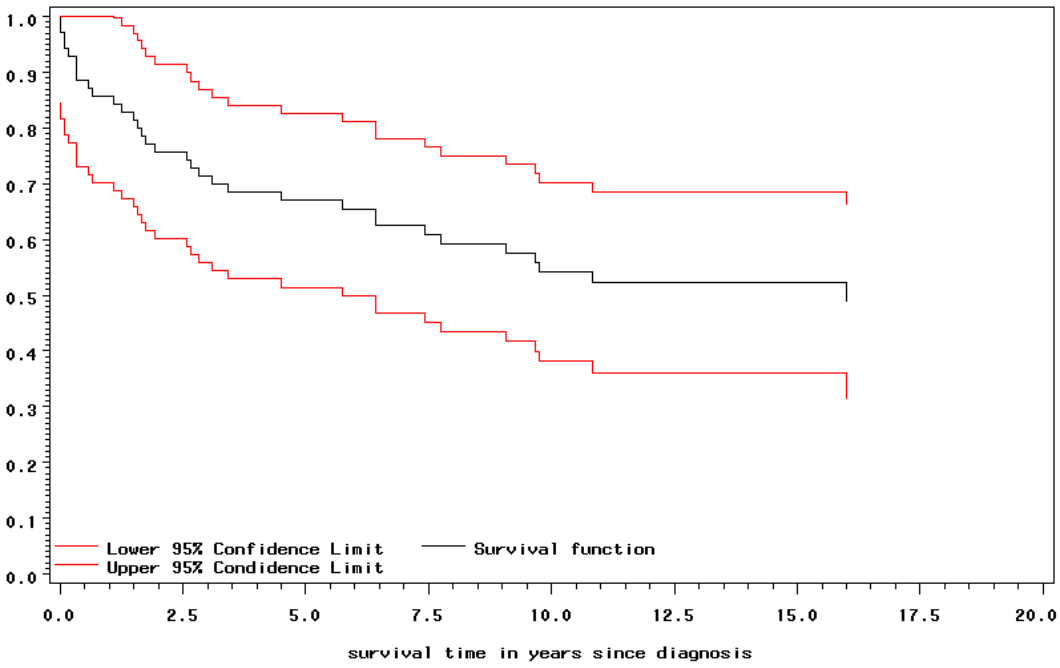


Figure 4.5: Hall-Wellner Confidence Bands for the Refused Group

Wilcoxon test (for right-censored data) to investigate if this difference is significant. Let F_C and F_S be the respective distribution function of X for the control group and the study group. The hypothesis to be tested is $H_0 : F_C = F_S$ versus the alternative $H_A : H_0$ *is not true*. The Wilcoxon test statistic has an asymptotic chi-square distribution with 1 degree of freedom.

The Wilcoxon test yields a test statistic of 17.4663 which gives a p-value of < 0.0001. Thus the distribution function of X of the control group is significantly different at the 5% level from that of the study group.

Similar test showed no significant difference in distributions between the control group and the refused group. The test statistic is 0.2610 which gives a p-value of 0.6095.

However, the distribution function of X of the study group is significantly different (at the 5% level) from that of the refused group. The test statistic is 4.6019 which yields a p-value of 0.0319.

4.2.2 Kaplan-Meier Estimate of Survival Function by Stage

The higher survival probability in the study group than that of the control group can be attributed to the fact that the study group has a higher proportion of early detection than that of the control group. This is studied in Chapter 5.

We would like to see how the Kaplan-Meier estimate of the survival function estimate changes when the data is analyzed by detection stage. It can be seen

in Table 5.1 that there are only a handful of women diagnosed beyond stage II. Because the sample size of the stages beyond II is too small for statistical analysis, the Kaplan-Meier estimate of the survival function will only be produced for the breast cancer cases diagnosed in stage I and stage II. As mentioned in Section 2.3, clinical stage I or II detection cases were combined with stage I detection cases. For simplicity, the combination of stage I detection and clinical stage I or II detection is called "combined stage I" detection.

The survival function for combined stage I detected patients is given in Figure 4.6 and the censoring pattern is given in Figure 4.7. In the refused group, there are only 2 breast cancer deaths, thus the survival curve is flat from time 2.5 years on. The confidence bands for the survival functions of combined stage I detection for the control group and the study group are given in Figure 4.8 and Figure 4.9.

Figure 4.6 clearly shows that for patients detected at combined stage I the study group has a better survival probability than that of the control group, especially in the first 10 years. A generalized Wilcoxon test is conducted to compare the distribution function of X of the control group and the study group. The test turns out to be not significant. The test statistic is 2.4668 which yields a p-value of 0.1163.

For patients detected at stage II, the survival function is given in Figure 4.10, while the censoring pattern is given in Figure 4.11. The confidence bands for the

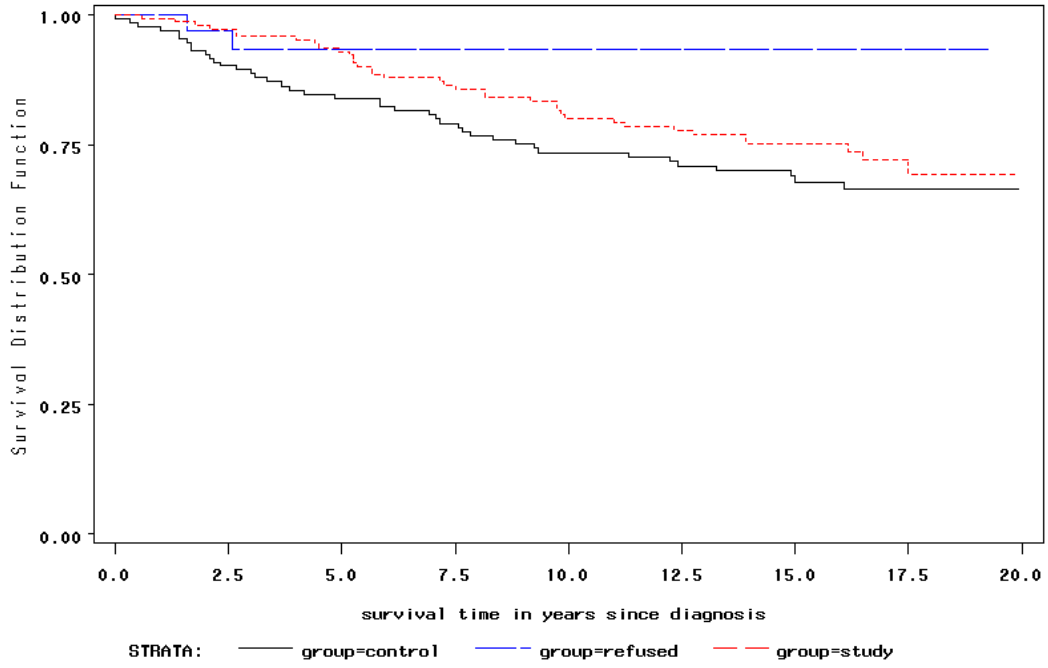


Figure 4.6: Kaplan-Meier Estimates for Combined Stage I Detection

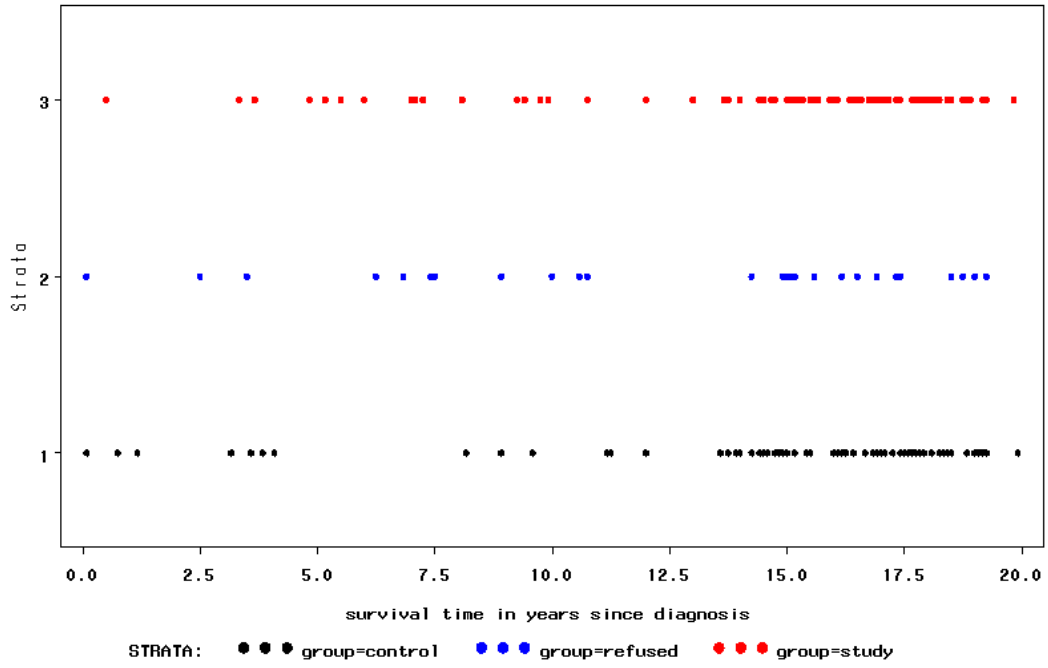


Figure 4.7: Censoring Pattern for Combined Stage I Detection

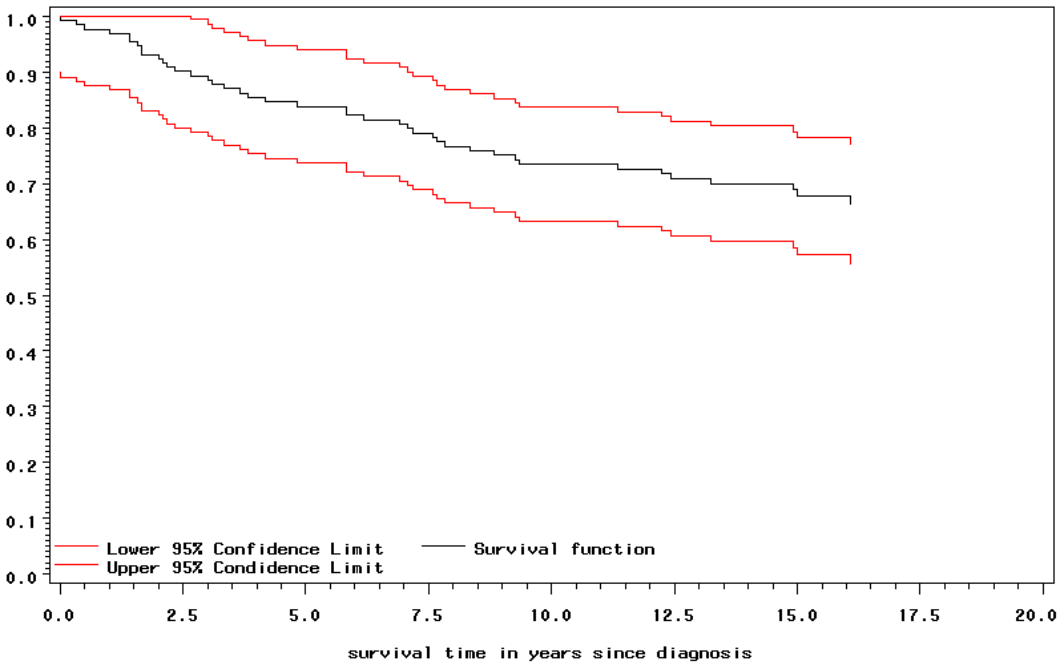


Figure 4.8: Hall-Wellner Confidence Bands for Combined Stage I Detection in the Control Group

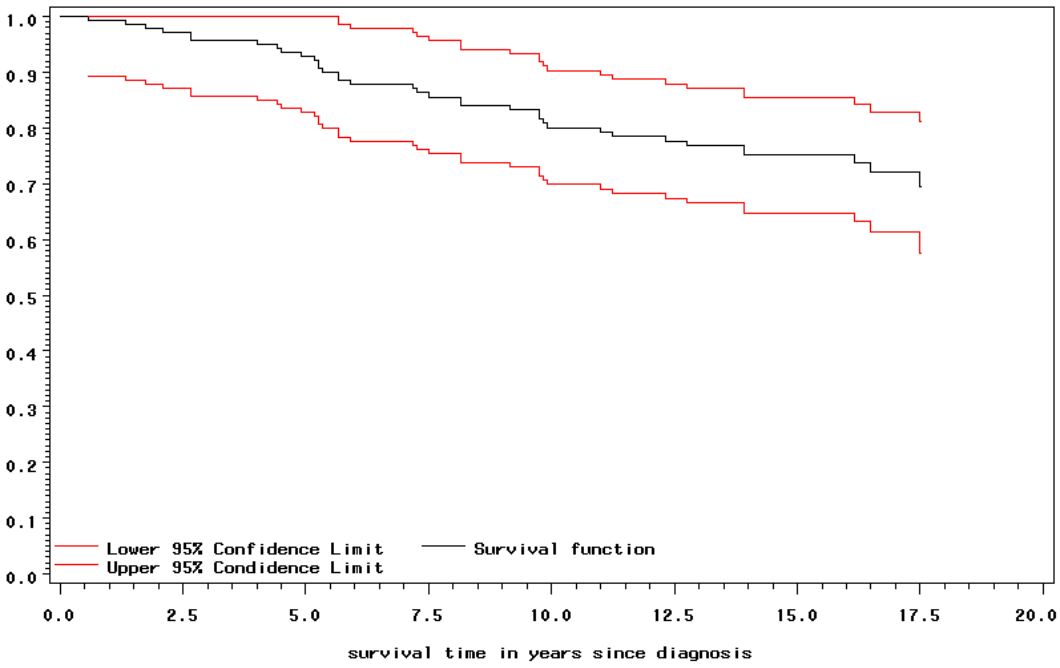


Figure 4.9: Hall-Wellner Confidence Bands for Combined Stage I Detection in the Study Group

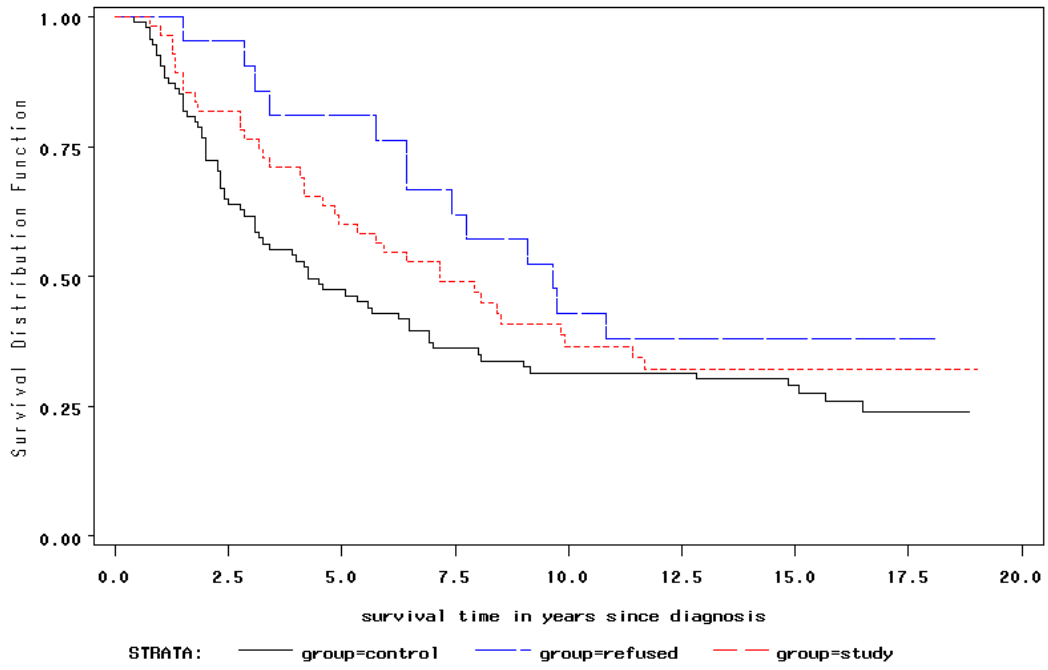


Figure 4.10: Kaplan-Meier Estimates for Stage II Detection

survival functions of stage II detection for the control group and the study group are given in Figure 4.12 and Figure 4.13.

Comparing Figure 4.6 and Figure 4.10, it can be easily seen that combined stage I detection has a much better survival probability than that of stage II detection. Although the study group has better survival curves in both stage I and II, than that of the control group, formal hypothesis testing showed no significant difference in the distribution function of X in the study and control groups. As in stage I detection, the test statistic is 2.7855 which gives a p-value of 0.0951.

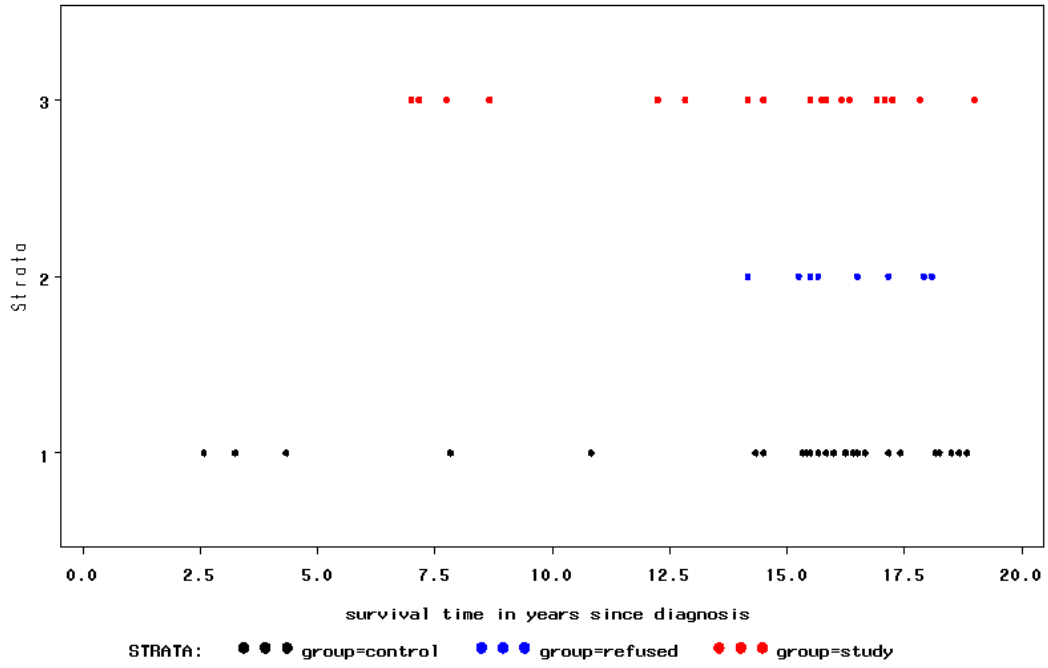


Figure 4.11: Censoring Pattern for Stage II Detection

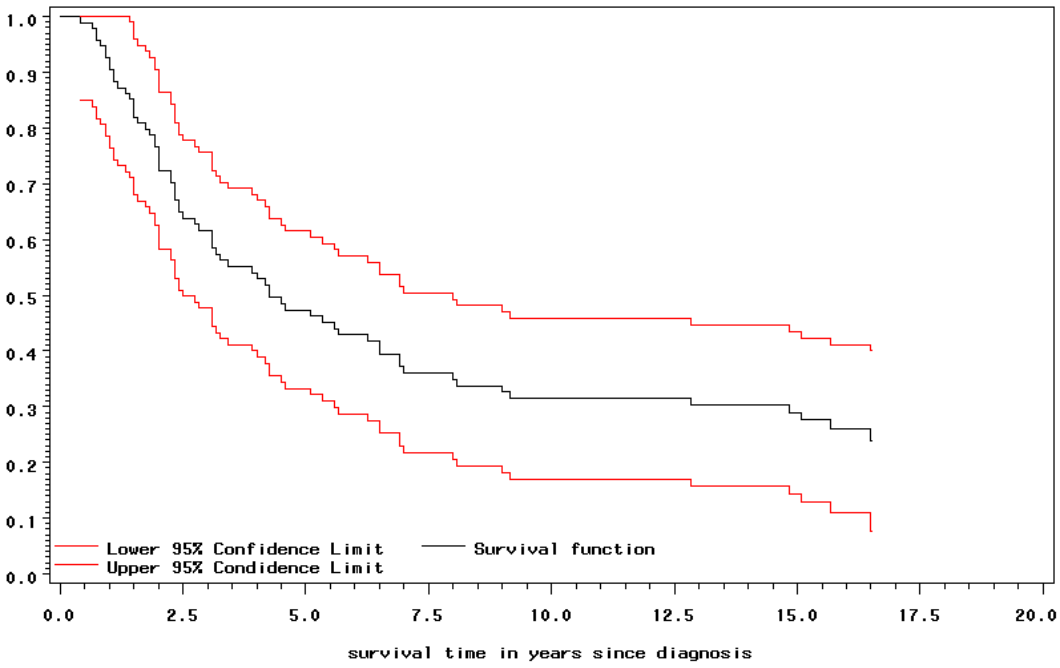


Figure 4.12: Hall-Wellner Confidence Bands for Stage II Detection - the Control Group

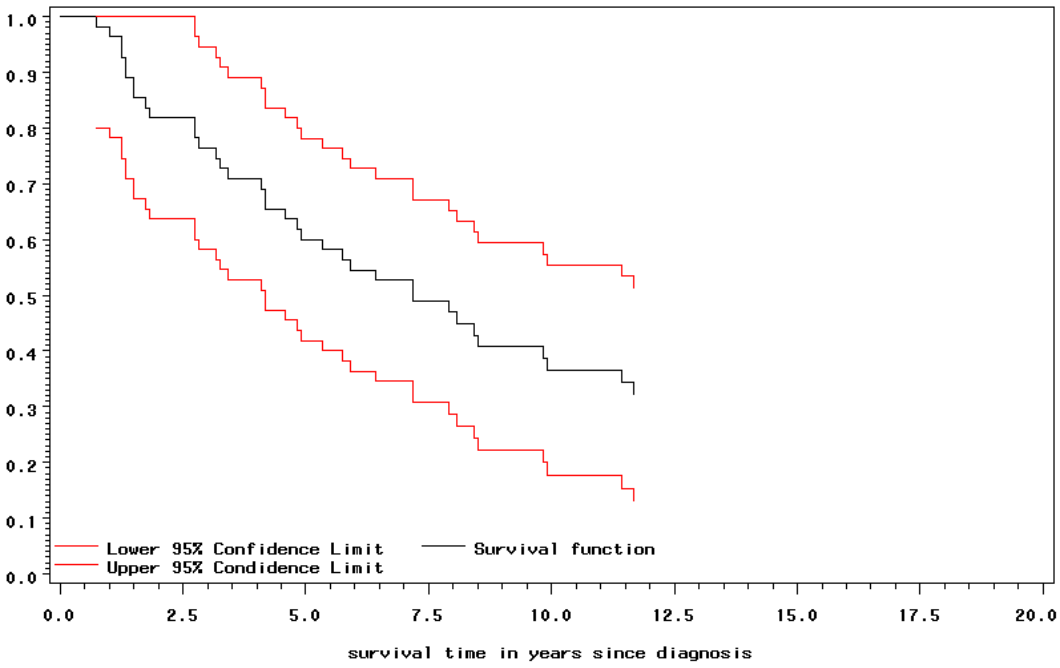


Figure 4.13: Hall-Wellner Confidence Bands for Stage II Detection - the Study Group

Chapter 5

Analysis of the Stage of Breast Cancer at Detection

The screening program is conducted on the assumption that breast cancer will be diagnosed at early stages of the disease development. Thus we would expect a larger proportion of early detection of breast cancer cases in the study group than in the control group. Table 5.1 gives the proportion of cancer detections by stage in the control group, the study group and the refused group. It is interesting to note that 63.98% of the cancer cases in the study group were detected at stage I, in contrast to 44.78% and 39.44% of the cancer cases detected in the control group and the refused group, respectively. This supports the belief that screening results in early detection of breast cancer. Formal statistical analysis will be conducted in subsequent sections.

However, it is surprising to see that some cancer cases in the study group were diagnosed in much later stages, that is, stage III or stage IV. Nonetheless, their percentages were small. There are several possible explanations for this. One possible explanation is that the breast cancer screening missed the cancer

Table 5.1: Breast Cancer Detection by Stage

Stage	Control		Study		Refused	
	N	%	N	%	N	%
Stage I	120	44.78%	135	63.98%	28	39.44%
Stage II	94	35.07%	55	26.07%	21	29.58%
Stage III	29	10.82%	9	4.27%	6	8.45%
Stage IV	6	2.24%	3	1.42%	10	14.08%
Clinical Stage I or II	14	5.22%	8	3.79%	4	5.63%
Unknown	5	1.87%	1	0.47%	2	2.82%
Total	268		211		71	

N = no. of cases, % = percentage

(false negative). Another is that some of the breast cancer cases diagnosed at the initial screening were already at advanced stages. In fact, in the study group 44.44% (4 out of 9) of the stage III detections were diagnosed either at initial screening or on early recall of the initial screening, while 33.33% (1 out of 3) of the stage IV detection and 50% (4 out of 8) of the clinical stage I or II detection were also diagnosed at initial screening.

5.1 Comparing the Distribution of Detection Stages in the Three Groups

We perform statistical hypothesis testing to check if the distribution of the detection stages are significantly different and if screening indeed leads to early detection of breast cancer. By early detection, it is meant to detect in stage I or II. Since there were only a few cancer cases detected in some advanced stages, for hypothesis testing we combine more advanced stages into one to increase the sample sizes. We shall call this combined category the advanced stages. Since our primary interest is in early detection, combining more advanced stages will not affect our investigation. As mentioned in Section 2.3, clinical stage I or II detection cases were combined with stage I detection cases. For simplicity, the combination of stage I detection and clinical stage I or II detection is called "combined stage I". Table 5.2 reflects these changes.

Table 5.2: Breast Cancer Detection by Stage – Collapsed Cells

Stage	Control		Study		Refused	
	N	%	N	%	N	%
Combined Stage I	134	50.00%	143	67.77%	32	45.07%
Stage II	94	35.07%	55	26.07%	21	29.58%
Advanced stages	40	14.93%	13	6.16%	18	25.35%
Total	268		211		71	

N = no. of cases, % = percentage

We first compare all three groups. Let p_{CS_1-C}, p_{S_2-C} and p_{adv-C} be the true probabilities of detections in combined stage I, stage II and advanced stages in the control group. Let p_{CS_1-S}, p_{S_2-S}, p_{adv-S}, p_{CS_1-R}, p_{S_2-R} and p_{adv-R} be similarly defined for the study group and the refused group. The hypothesis to be tested is $H_0 : p_{CS_1-C} = p_{CS_1-S} = p_{CS_1-R},\ p_{S_2-C} = p_{S_2-S} = p_{S_2-R},\ p_{adv-C} = p_{adv-S} = p_{adv-R}$ versus the alternative $H_A : H_0$ *is not true*. The chi-square statistic yields 28.3772 with 4 degrees of freedom which gives a p-value of < 0.0001. Thus we reject the null hypothesis of equal proportions at the 5% level.

However, this test does not give us a sense as to which pairs are different. Thus we conduct a pairwise comparison with a smaller level of significance, that is 1.67% level (5% divided by 3).

We first compare the control group and the study group. The hypothesis to be tested is $H_0 : p_{CS_1-C} = p_{CS_1-S},\ p_{S_2-C} = p_{S_2-S},\ p_{adv-C} = p_{adv-S}$ versus

the alternative $H_A : H_0$ *is not true*. For the hypothesis H_0 of equal multinomial distributions of the detection probabilities, the chi-square statistic is 17.7233 with 2 degrees of freedom, which gives a p-value 0.0001. Thus, we conclude that the detection stage distributions in the control group and the study group are significantly different at the 1.67% level.

A similar test also shows a significant difference between the study group and the refused group at the 1.67% level. The test yields a chi-square statistic of 22.4531 with 2 degrees of freedom which gives a p-value < 0.0001.

But the test shows no significant difference in the detection stage distribution between the control group and the refused group. The chi-square statistic is 4.3452 with 2 degrees of freedom which gives a p-value of 0.1139.

Next we investigate differences in the probability of the combined stage I detection among the three groups. The sample proportion of combined stage I detection in the control group, the study group and the refused group are $\hat{p}_{S_1-C} = 0.5000$, $\hat{p}_{S_1-S} = 0.6777$ and $\hat{p}_{S_1-R} = 0.4507$ respectively; see Table 5.2.

Comparison of the control group and the study group shows a significant difference in the probability of combined stage I detection (the chi-square statistic is 15.2904 with 1 degree of freedom which gives a p-value of < 0.0001). The sample proportion of combined stage I detection in the study group is 0.1777 higher than that in the control group. The standard error of the difference of the sample proportion is 0.0444 which gives an asymptotic 95% confidence interval of

$\hat{p}_{CS_1-S} - \hat{p}_{CS_1-C}$ (0.0908, 0.2647). Furthermore, the Mantel-Haenszel relative risk for these two groups is 1.5515 with a 95% confidence interval (1.2334, 1.9515). This means that the breast cancer cases in the study group are 55% more likely to be diagnosed at stage I or clinical stage I or II than those in the control group.

The difference between the study group and the refused group is also significant (the chi-statistic is 11.6279 with 1 degree of freedom which gives a p-value of 0.0006). The sample proportion of combined stage I in the study group is 0.2270 higher than that in the refused group with a standard error of the difference of the two sample proportions equal to 0.0672 and an asymptotic 95% confidence interval of (0.0952 0.3588) for the difference $\hat{p}_{CS_1-S} - \hat{p}_{CS_1-R}$. The Mantel-Haenszel relative risk for these two groups is 1.7044 with a 95% confidence interval of (1.2785, 2.2723). Thus the breast cancer cases in the study group are 70% more likely to be diagnosed in stage I or clinical stage I or II than those cases in the refused group.

The difference between the control group and the refused group is not significant.

5.2 Analysis by Outcomes at the End of the HIP Study and Total Years Lived

As discussed in previous sections, at the end of the follow-up period each patient has one of the three possible outcomes: survival, death from breast cancer or death from other causes. We have investigated in Chapter 4 the differences in outcomes among the three groups. In this section, we carry out a more refined study by examining the effect of early detection on the outcomes. Table 5.3 contains the sample proportions for each possible outcome in each particular stage.

We shall use the data in Table 5.3 to compare the outcomes the patients whose cancer were detected in combined stage I or stage II among the three groups. These stages are deemed as early detections. Since only few cases were detected in more advanced stages and since they are not our main interest, we will not carry out the statistical analysis of cases.

For cancer detected in combined stage I, there was no significant difference in the outcome probabilities between the control group and the study group. The chi-square statistic is 1.7194 with 2 degrees of freedom which gives a p-value of 0.4233.

Table 5.3: Breast Cancer Detection Stage by Outcomes at the End of the HIP Study

Stage and Group		Survival		DBC		DOC	
		N	%	N	%	N	%
	Control	75	55.97%	41	30.60%	18	13.43%
Combined	Study	81	56.64%	36	25.17%	26	18.18%
Stage I	Refused	18	56.25%	2	6.25%	12	37.50%
	Control	21	22.34%	68	72.34%	5	5.32%
Stage II	Study	12	21.82%	36	65.45%	7	12.73%
	Refused	5	23.81%	13	61.90%	3	14.29%
	Control	1	2.50%	38	95.00%	1	2.50%
Advanced	Study	0	0.00%	12	92.31%	1	7.69%
Stages	Refused	0	0.00%	18	100.00%	0	0.00%

N = no. of cases, % = percentage

DBC = death from breast cancer, DOC = death from other causes

Also, there is no significant difference in outcome probabilities for cancer detected in stage II between the control group and the study group. The chi-square statistics is 2.6044 with 2 degree of freedom which give a p-value 0.2719.

The refused group has too few observations to carry out the statistical comparison with the two other groups.

It is interesting to note that for advanced stages detection, the proportions of survival and deaths due to other causes are extremely small.

Table 5.4 contains the sample average and standard deviation of total years lived after detection for all cancer cases broken down by detection stage. Table 5.5 contains the same information as in Table 5.4, only the information is further broken down by the outcomes at the end of HIP study.

From inspection of Table 5.4, there seems to be little difference among groups in the average of total years lived after detection for all cancer cases detected in combined stage I. While for stage II and advanced stages detection, the average of total years lived after detection in the control group is very close to that in the study group. However, the refused group has a higher average for stage II detection, and a lower average for advanced stages detection.

When the total years lived after detection is further broken down by outcomes at the end of HIP study (Table 5.5), there is little difference in the average of total years lived after detection for women who survived breast cancer regardless the stage detection. We expect this because the total years lived for women who

Table 5.4: Average and Standard Deviation of Total Years Lived after Detection by Breast Cancer Detection Stage

Stage and Group		All Cases		
		N	Average	Std Dev
Combined Stage I	Control	134	12.13	6.100
	Study	143	13.10	5.166
	Refused	32	12.30	5.854
Stage II	Control	94	7.00	6.219
	Study	55	8.00	5.586
	Refused	21	10.20	5.478
Advanced Stages	Control	40	3.20	4.005
	Study	13	2.37	2.829
	Refused	18	1.86	3.716

N = no. of cases, Std Dev = standard deviation

Table 5.5: Average and Standard Deviation of Total Years Lived after Detection by Cancer Detection Stage – Continued

Stage and Group		Survival			DBC			DOC		
		N	Avg	SD	N	Avg	SD	N	Avg	SD
	Control	75	16.60	1.651	41	5.85	4.478	18	7.76	5.333
Combined	Study	81	16.66	1.468	36	7.75	4.474	26	9.44	4.534
Stage I	Refused	18	16.67	1.616	2	2.08	0.707	12	7.44	4.074
	Control	21	16.68	1.362	68	4.10	3.796	5	5.77	3.481
Stage II	Study	12	16.39	1.348	36	4.78	3.147	7	10.17	3.342
	Refused	5	16.50	1.174	13	6.46	3.009	3	15.92	1.887
	Control	1	17.92	.	38	2.61	2.995	1	11.17	.
Advanced	Study	0	.	.	12	1.69	1.446	1	10.58	.
Stages	Refused	0	.	.	18	1.86	3.716	0	.	.

N = no. of cases, Avg = average, SD = standard deviation

DBC = death from breast cancer, DOC = death from other causes

survived breast cancer are censored by end of follow-up. As for women who died from breast cancer, women diagnosed at combined stage I in the study group have a longer average of total years lived after detection than in the control group. The refused group only has 2 observations and is therefore excluded from comparison. For stage II detection, the averages of women who died from breast cancer are similar in the control group and the study group, while the refused group seems to have a higher average than the other groups. For advanced stage detection, the study group and the refused group have a similar average, while the control group has a higher average compare to the other groups.

Table 5.6 contains the average and standard deviation of total years lived after birth, in years, for all cancer cases broken down by stage at detection. Table 5.7 contains the same information as in Table 5.6, only the information is broken down further by outcomes at the end of HIP study.

No formal test is conducted to see if there is a significant difference in the average of total years lived after birth among the control group, the study group and the refused group. However, from inspection it can be seen that the study and control groups have similar averages of total years lived after birth in all stages of detection.

When the average of total years lived after birth is further broken down by outcomes at the end of the HIP study, for combined stage I detection, the study group always has either similar or longer average as compare with the other two

Table 5.6: Average and Standard Deviation of Total Years Lived after Birth by Cancer Detection Stage

Stage and Group		All Cases		
		N	Average	Std Dev
Combined Stage I	Control	134	67.59	9.119
	Study	143	68.21	8.306
	Refused	32	66.29	7.363
Stage II	Control	94	61.42	8.068
	Study	55	62.60	8.565
	Refused	21	64.36	7.917
Advanced Stages	Control	40	58.24	7.574
	Study	13	56.28	8.292
	Refused	18	57.08	6.077

N = no. of cases, Std Dev = standard deviation

Table 5.7: Average and Standard Deviation of Total Years Lived after Birth by Cancer Detection Stage – Continued

Stage and Group		Survival			DBC			DOC		
		N	Avg	SD	N	Avg	SD	N	Avg	SD
	Control	75	72.20	6.169	41	60.27	8.712	18	65.08	8.754
Combined	Study	81	71.39	6.315	36	62.62	8.788	26	66.07	8.577
Stage I	Refused	18	69.17	5.080	2	54.08	3.653	12	64.00	8.083
	Control	21	69.91	6.887	68	58.50	6.462	5	65.47	5.778
Stage II	Study	12	69.38	6.801	36	59.25	7.482	7	68.24	7.248
	Refused	5	69.57	6.952	13	61.52	7.863	3	67.97	4.394
	Control	1	66.33	.	38	57.62	7.192	1	73.83	.
Advanced	Study	0	.	.	12	55.04	7.293	1	71.17	.
Stages	Refused	0	.	.	18	57.08	6.077	0	.	.

N = no. of cases, Avg = average, SD = standard deviation

DBC = death from breast cancer, DOC = death from other causes

groups. While the refused group has the lowest average among the three groups for combined stage I detection. For stage II detection, the average of the total years lived since birth for women who survived breast cancer is similar in all three groups.

All in all, we do not see many difference among the three groups in the averages of total years lived whether it is measured from birth or from the time of cancer detection.

Chapter 6

Analysis of Five-Year Survival Probability and Normalized Data

Ideally, studies of the efficacy of screening should include information about the curing probability of the breast cancer patients. Curing means that a patient has completed the treatment and the doctor has declared that she is free of breast cancer or her cancer is in remission. It will also be relevant to know the recurrence probability of the breast cancer among women who were previously declared to be free of cancer. However, the HIP study did not collect such information. In the absence of such data, we will investigate the five-year survival probability instead.

In this chapter we also will compare the un-normalized and normalized yearly proportion of women who survived breast cancer, died from breast cancer and died from other cause. By conducting this analysis, we hope to learn about the conditional survival probability and breast cancer mortality.

6.1 Five-year Survival Probability

The five-year survival probability is generally used as a measure to assess the success of a cancer treatment. The five-year survival is measured from the time of cancer diagnosis. Those women who survive five years can be in either a cancer free state, or a therapy state or a recurrence state.

Some medical researchers argues that for the breast cancer, five years is not long enough to follow the history of the disease. They proposed a ten years or even fifteen years of follow-up. The HIP study had about 3.5 years of screening (although the actual screening period was longer than that) and 15 years of follow-up. Thus we have about 18 years of history of the breast cancer development. It enables us to analyze a ten-year survival probability as well. We will also analyze a shorter three-year survival probability for comparison purposes.

Table 6.1, Table 6.2 and Table 6.3 give a yearly account on how many women survived breast cancer, died from breast cancer and died from other causes in each of the three groups. These numbers are calculated based on the breast cancer diagnosed date, not the entry date to the HIP study.

Yearly survival proportions for each possible outcome (survived the breast cancer, died from breast cancer or died from other causes) are given in Figure 6.1, Figure 6.2 and Figure 6.3. Each graph consists of three lines, one for each group. The proportion is computed as the number of occurrences of a particular outcome divided by the total number of cases.

Table 6.1: Yearly Data of Women Survived Breast Cancer, Died From Breast Cancer and Died from Other Causes among Cancer Cases in the Control Group

End of n^{th} Year	Survival	Death from Breast Cancer	Death from Other Causes	Total Cases
1	241	25	2	268
2	206	59	3	268
3	186	78	4	268
4	168	91	9	268
5	155	102	11	268
6	147	110	11	268
7	139	118	11	268
8	132	124	12	268
9	125	129	14	268
10	120	133	15	268
11	118	134	16	268
12	113	136	19	268
13	109	139	20	268
14	106	140	22	268
15	104	142	22	268
16	100	145	23	268
17	97	147	24	268
18	97	147	24	268

Table 6.2: Yearly Data of Women Survived Breast Cancer, Died from Breast Cancer and Died from Other Causes among Cancer Cases in the Study Group

End of n^{th} Year	Survival	Death from Breast Cancer	Death from Other Causes	Total Cases
1	203	7	1	211
2	190	20	1	211
3	181	29	1	211
4	175	33	3	211
5	163	44	4	211
6	150	54	7	211
7	147	55	9	211
8	136	61	14	211
9	129	66	16	211
10	118	73	20	211
11	116	73	22	211
12	112	77	22	211
13	107	79	25	211
14	103	81	27	211
15	101	81	29	211
16	98	81	32	211
17	94	83	34	211
18	93	84	34	211

Table 6.3: Yearly Data of Women Survived Breast Cancer, Died from Breast Cancer and Died from Other Causes among Cancer Cases in the Refused Group

End of n^{th} Year	Survival	Death from Breast Cancer	Death from Other Causes	Total Cases
1	60	10	1	71
2	53	17	1	71
3	49	20	2	71
4	46	22	3	71
5	45	23	3	71
6	44	24	3	71
7	40	26	5	71
8	36	28	7	71
9	35	28	8	71
10	31	31	9	71
11	28	32	11	71
12	28	32	11	71
13	28	32	11	71
14	28	32	11	71
15	26	32	13	71
16	25	32	14	71
17	24	33	14	71
18	23	33	15	71

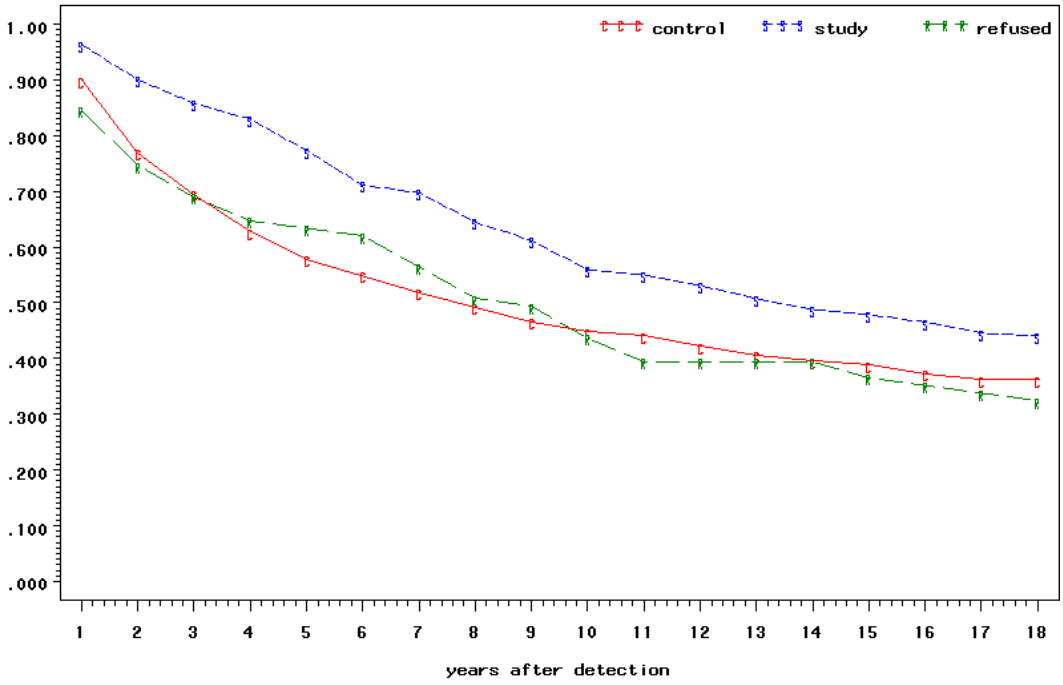


Figure 6.1: Proportion of Survival

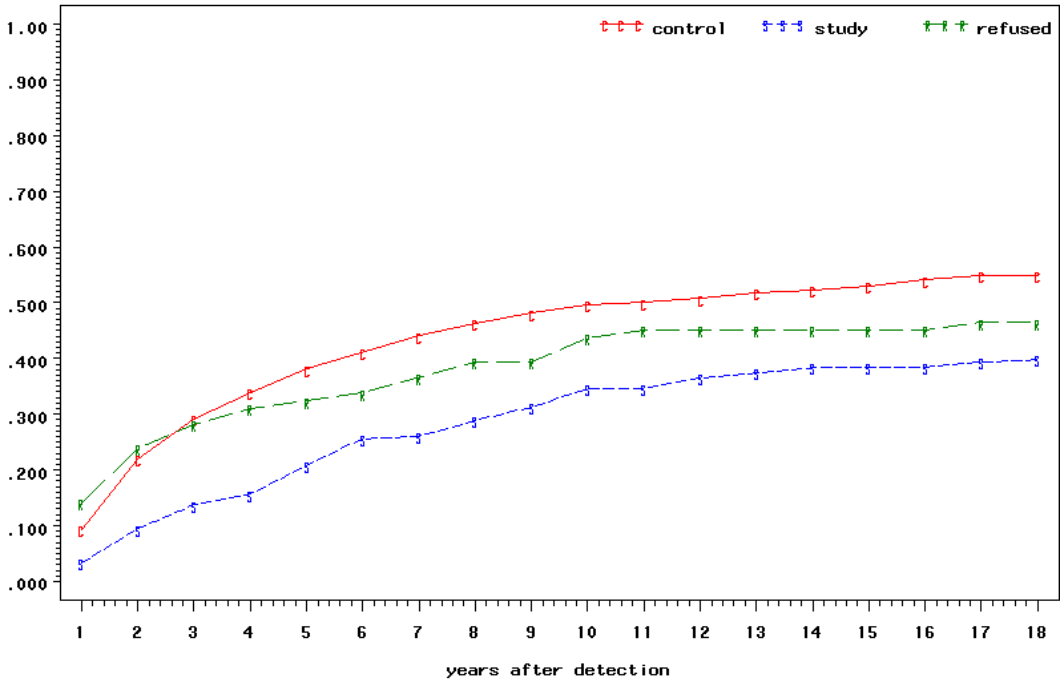


Figure 6.2: Proportion of Death from Breast Cancer

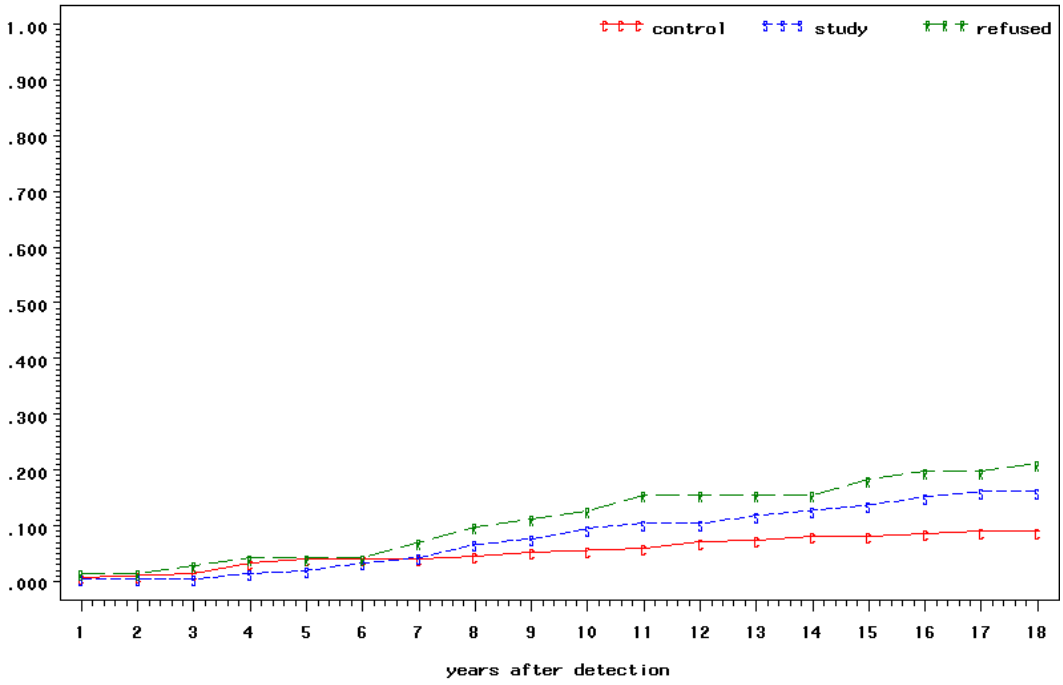


Figure 6.3: Proportion of Death from Other Causes

Table 6.4: Proportions of Survival, Deaths from Breast Cancer and Death from Other Causes Three Years after Diagnosis

	Proportions of		
Group	Survival	DBC	DOC
Control	69.4%	29.1%	1.5%
Study	85.8%	13.7%	0.5%
Refused	69.0%	28.2%	2.8%

DBC = Deaths from Breast Cancer

DOC = Deaths from Other Causes

The graphs show that the survival proportion of the study group is always higher than that of the control group. Also the proportion of deaths from breast cancer in the study group is always lower than that of the control group.

As for the refused group, the graphs do not show a consistent pattern. Sometimes it is similar to the control group and some other times it is close to that of the study group.

The survival proportions, deaths from breast cancer and deaths from other causes of the three groups three years after diagnosis where given in Table 6.4. While Table 6.5 and Table 6.6 contain similar information only for five and ten years after diagnosis respectively.

It can be seen from Figure 6.1 that the largest survival proportion difference between the study group and the control group happens in the fourth and

Table 6.5: Proportions of Survival, Deaths from Breast Cancer and Death from Other Causes Five Years after Diagnosis

	Proportions of		
Group	Survival	DBC	DOC
Control	57.8%	38.1%	4.1%
Study	77.3%	20.9%	1.9%
Refused	63.4%	32.4%	4.2%

DBC = Deaths from Breast Cancer

DOC = Deaths from Other Causes

Table 6.6: Proportions of Survival, Deaths from Breast Cancer and Death from Other Causes Ten Years after Diagnosis

	Proportions of		
Group	Survival	DBC	DOC
Control	44.8%	49.6%	5.6%
Study	55.9%	34.6%	9.5%
Refused	43.7%	43.7%	12.7%

DBC = Deaths from Breast Cancer

DOC = Deaths from Other Causes

fifth year after diagnosed. After the fifth year, the difference decreases as time increases.

It is also interesting to note that the proportion of deaths from other causes in the study group and the refused group increases at a faster rate over time than that in the control group. As discussed in Chapter 3, this probably happened because more women survived breast cancer in the study group but then they died from other causes. Unfortunately, the HIP study has no specific record of the causes of death other than the breast cancer. If it had, it would have been possible to ascertain whether or not the other causes of death somehow is related to breast cancer.

We should however point out that the above is only an informal discussion. We have not carried out any formal statistical analysis such as variance calculation or hypothesis testing.

6.2 Normalized Data–Conditional Probabilities

Our next task is to compare changes in these proportions when the data are normalized for third, fifth and tenth year after detection. By normalizing, we mean using the number of survival of a particular year as the starting number and use this number as the denominator to compute the normalized proportion of three outcomes annually in the subsequent years. In other words, we are estimating the conditional probabilities of survival, death from breast cancer and

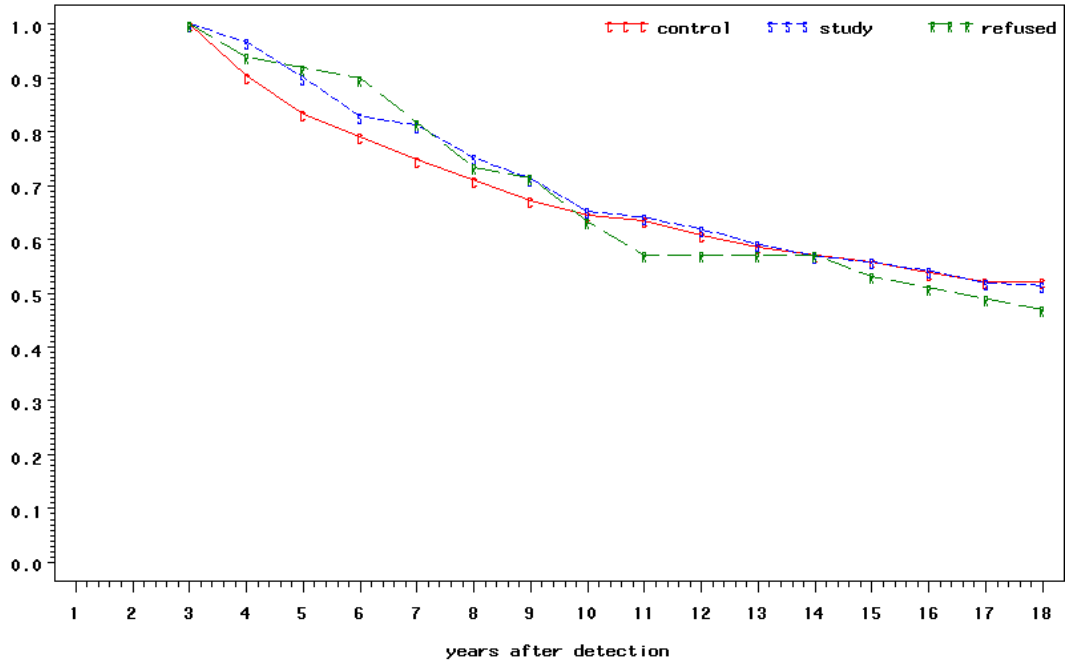


Figure 6.4: Proportion of Survival - Normalized for Year 3

death from other causes, given that a patient has survived a specified number of years, say five. The fifth-year normalized data for the control group, the study group and the refused group are given in Table 6.7, Table 6.8 and Table 6.9.

The graphs for the third year normalized proportion of each outcome are given in Figure 6.4, Figure 6.5 and Figure 6.6, for the fifth year are given in Figure 6.7, Figure 6.8 and Figure 6.9, and for the tenth year are given in Figure 6.10, Figure 6.11 and Figure 6.12.

In the first few years, third-year normalized survival proportions for the study group are higher than those of the control group before both proportions actually

Table 6.7: Fifth Year Normalized Data of Women Survived Breast Cancer, Died from Breast Cancer and Died from Other Causes in the Control Group

End of n^{th} Year	Survival	Death from Breast Cancer	Death from Other Causes	Total Cases
1	.	.	.	.
2	.	.	.	.
3	.	.	.	.
4	.	.	.	.
5	155	0	0	155
6	147	8	0	155
7	139	16	0	155
8	132	22	1	155
9	125	27	3	155
10	120	31	4	155
11	118	32	5	155
12	113	34	8	155
13	109	37	9	155
14	106	38	11	155
15	104	40	11	155
16	100	43	12	155
17	97	45	13	155
18	97	45	13	155

Table 6.8: Fifth Year Normalized Data of Women Survived Breast Cancer, Died from Breast Cancer and Died from Other Causes in the Study Group

End of n^{th} Year	Survival	Death from Breast Cancer	Death from Other Causes	Total Cases
1	.	.	.	.
2	.	.	.	.
3	.	.	.	.
4	.	.	.	.
5	163	0	0	163
6	150	10	3	163
7	147	11	5	163
8	136	17	10	163
9	129	22	12	163
10	118	29	16	163
11	116	29	18	163
12	112	33	18	163
13	107	35	21	163
14	103	37	23	163
15	101	37	25	163
16	98	37	28	163
17	94	39	30	163
18	93	40	30	163

Table 6.9: Fifth Year Normalized Data of Women Survived Breast Cancer, Died from Breast Cancer and Died from Other Causes in the Refused Group

End of n^{th} Year	Survival	Death from Breast Cancer	Death from Other Causes	Total Cases
1	.	.	.	.
2	.	.	.	.
3	.	.	.	.
4	.	.	.	.
5	45	0	0	45
6	44	1	0	45
7	40	3	2	45
8	36	5	4	45
9	35	5	5	45
10	31	8	6	45
11	28	9	8	45
12	28	9	8	45
13	28	9	8	45
14	28	9	8	45
15	26	9	10	45
16	25	9	11	45
17	24	10	11	45
18	23	10	12	45

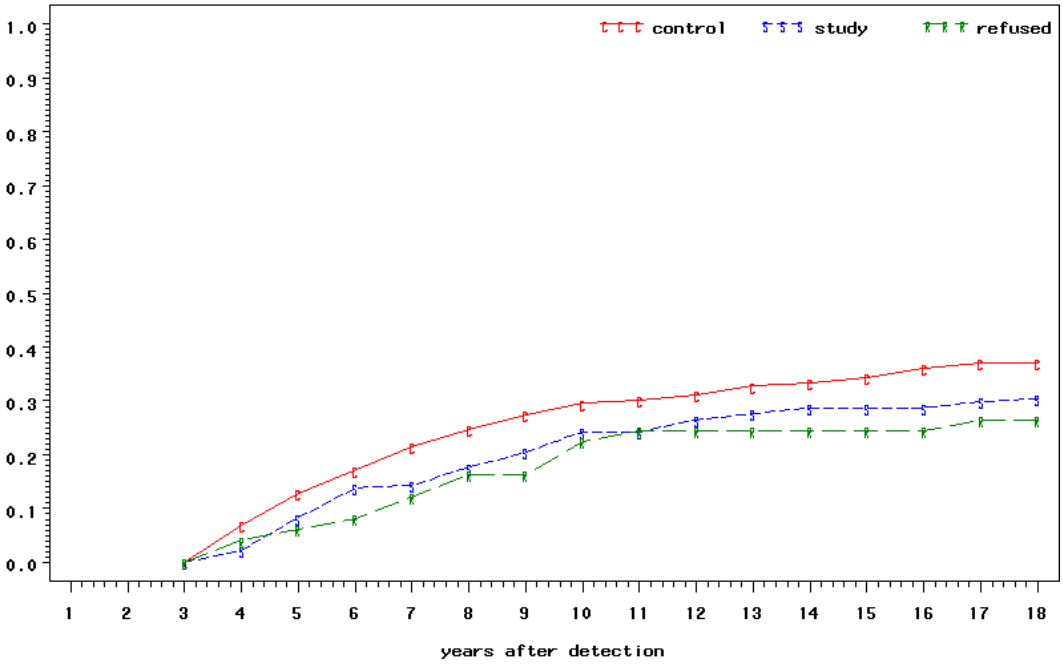


Figure 6.5: Proportion of Death from Breast Cancer - Normalized for Year 3

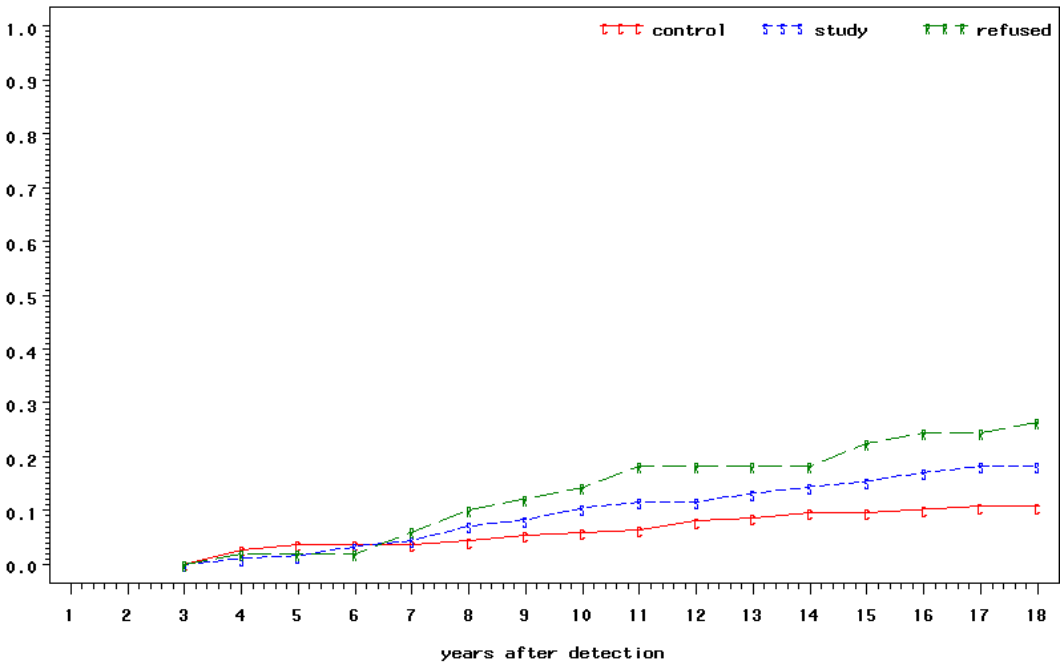


Figure 6.6: Proportion of Death from Other Causes - Normalized for Year 3

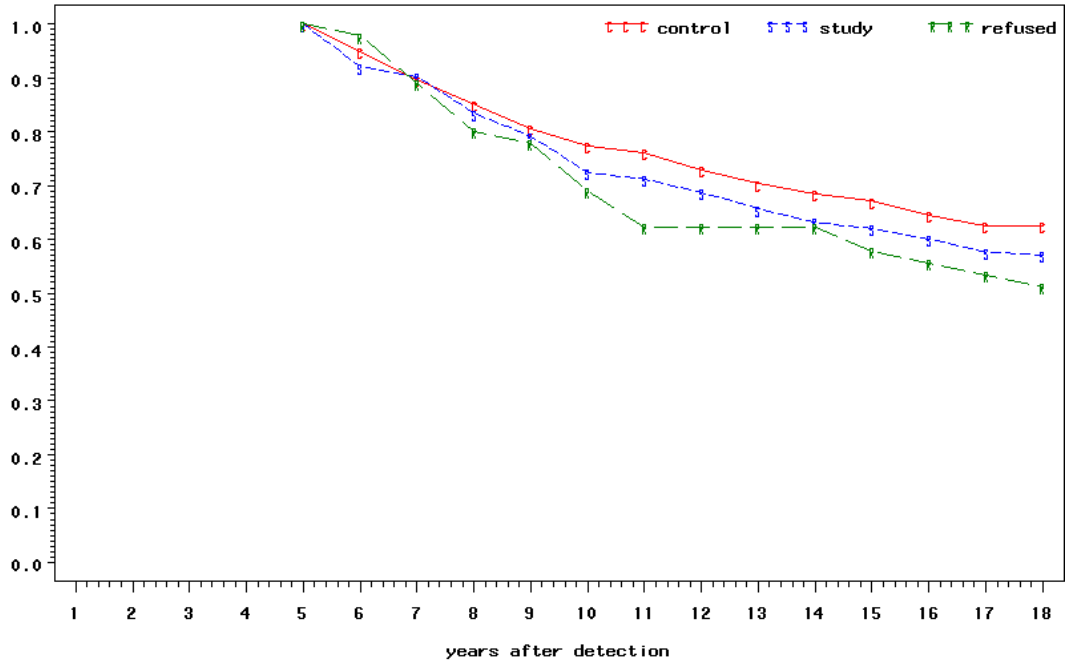


Figure 6.7: Proportion of Survival - Normalized for Year 5

converge. While the third-year normalized death from breast cancer proportions for the control group are higher than the study group. The third-year normalized death from other causes proportions in the study group are still higher than those of the control group although it seems that the differences are more pronounce compare to the un-normalized proportions.

For the fifth-year normalized survival proportion, there are several interesting facts. The fifth-year normalized survival curve of the control group actually is not better than that of the study group. Also, the cancer mortality curve of the control group is only slightly higher than those of the study group. However,

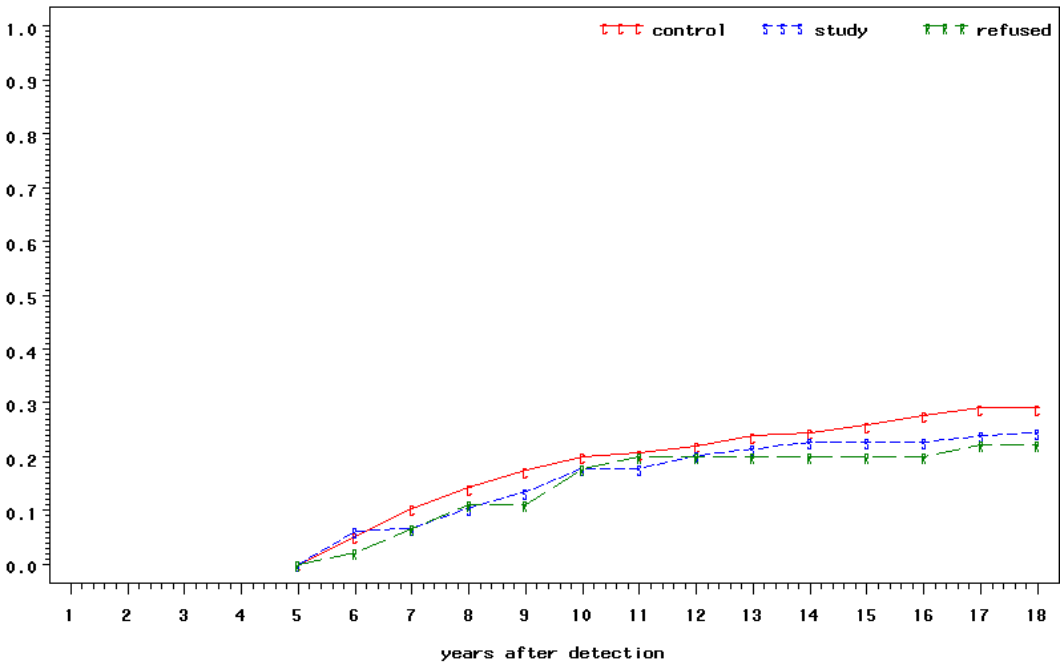


Figure 6.8: Proportion of Death from Breast Cancer - Normalized for Year 5

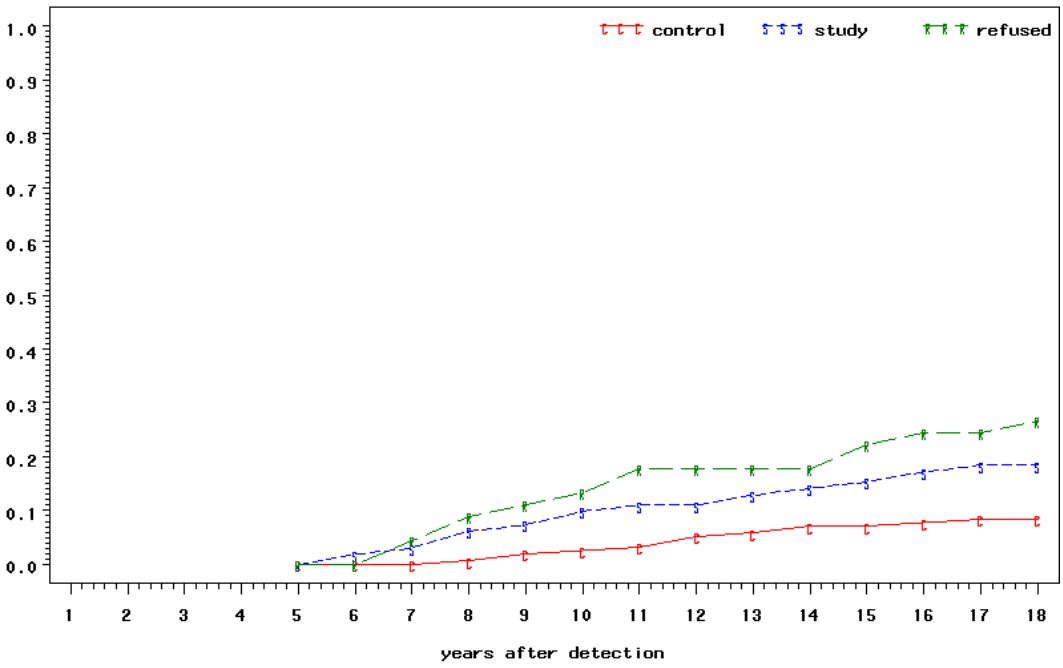


Figure 6.9: Proportion of Death from Other Causes - Normalized for Year 5

normalized curve of the death from other causes of the study group is much higher than that of the control group. Two conclusions can be drawn here. First, even though the fifth-year normalized survival proportions of the study group are not better than those of the control group, it does not mean that screening is not beneficial, since the fifth-year normalized proportion of deaths from other causes of the study group is still much higher than that of the control group. Second, if the breast cancer was not cured (or in remission) after five years, than it probably was either a deadly cancer or detected in an advanced stage for which screening offers no benefit and thus the fifth-year normalized curve of deaths from breast cancer are expected to be similar in the control group and the study group.

The tenth-year normalized proportions behave similarly to the fifth-year normalized proportions. The similarity is more pronounced as expected, since the effect of screening wear out as time goes on.

It will be interesting to conduct a refined study of the fifth-year and the tenth-year normalized data by detection stages.

The fifth-year normalized proportion of cancer by detection stage is given in Table 6.10. Comparing Table 6.10 to Table 5.1, that is subtracting number of cases for the fifth-year normalized data (Table 6.10) from number of cases for the un-normalized data (Table 5.1), gives the number of women died in the first five years after detection. Because there are only few observations beyond stage II, we shall only analyze data involving stage I and stage II detection. Table 6.11

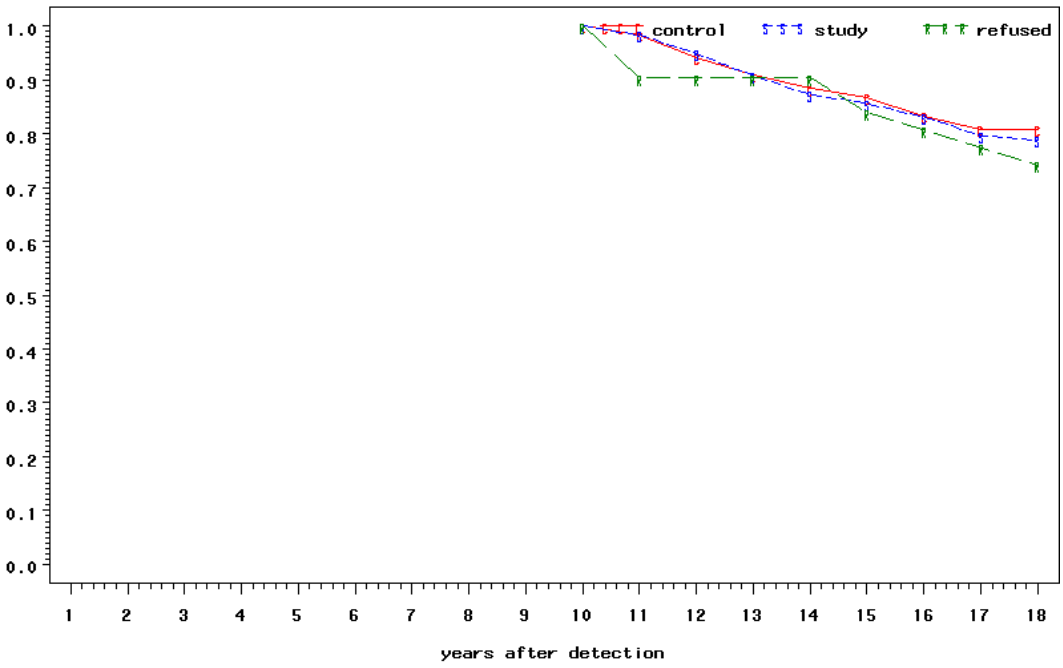


Figure 6.10: Proportion of Survival - Normalized for Year 10

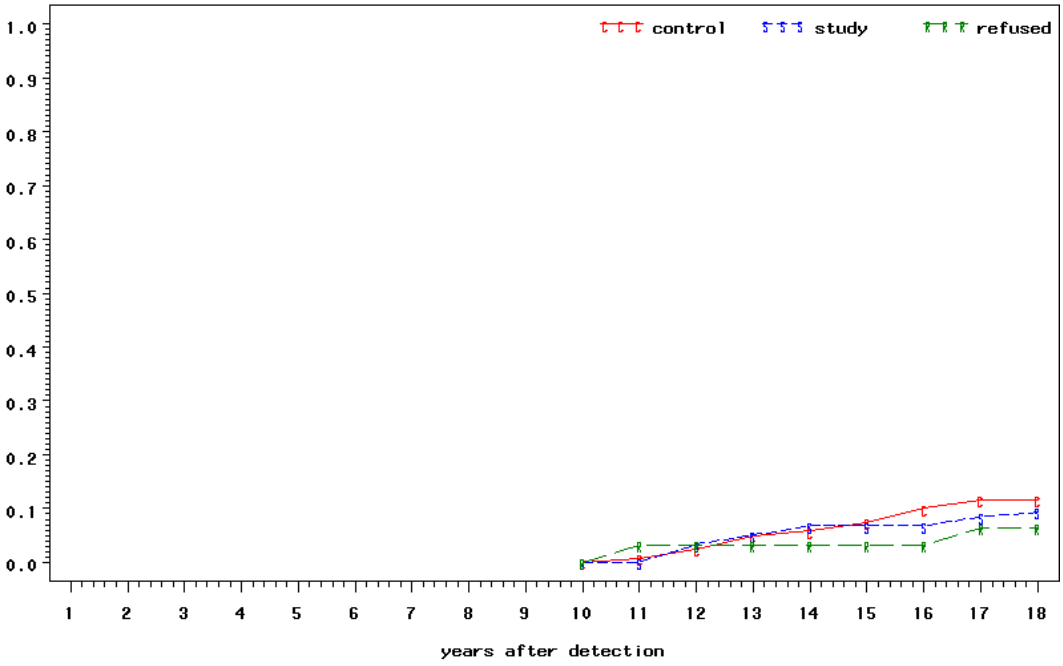


Figure 6.11: Proportion of Death from Breast Cancer - Normalized for Year 10

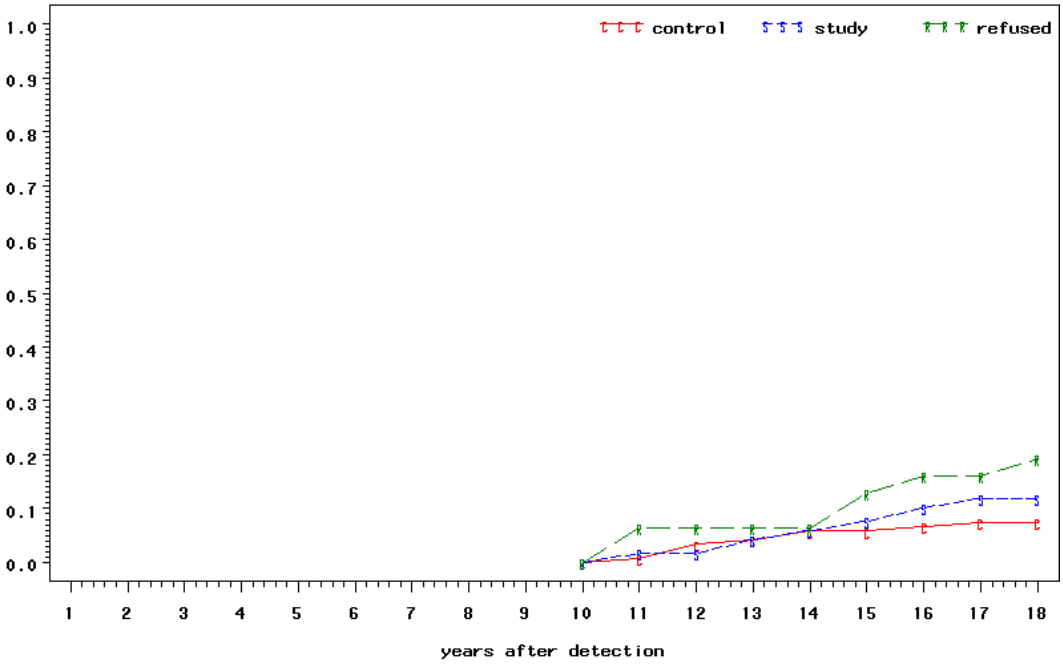


Figure 6.12: Proportion of Death from Other Causes - Normalized for Year 10

Table 6.10: Breast Cancer by Detection Stage - Normalized for Year 5

Stage	Control		Study		Refused	
	N	%	N	%	N	%
Stage I	93	60.00%	121	74.23%	23	51.11%
Stage II	42	27.10%	33	20.25%	17	37.78%
Stage III	8	5.16%	1	0.61%	1	2.22%
Stage IV	0	0.00%	0	0.00%	0	0.00%
Clinical Stage I or II	12	7.74%	8	4.91%	4	8.88%
Unknown	0	0.00%	0	0.00%	0	0.00%
Total	155		163		45	

N = no. of cases, % = percentage

contains the number of women died either from breast cancer or from other causes in the first five years after detection. The percentage was calculated with respect to number of cases of the un-normalized data.

From Table 6.11, we know that for stage I detection of the control group there is a total of 27 deaths in the first 5 years after detection, 20 of them or 16.67% of cancer cases in control group died from breast cancer. While for stage I detection of the study group for the same period of time, there is a total of 14 deaths, 10 of them or 7.41% died from breast cancer. Thus the proportion of cancer death in the control group is more than twice than the study group. However, for stage II detection, there is 52.13% of cancer death in the control group as opposed to

Table 6.11: Number of Death during the First Five Years after Diagnosis

Stage and Group		No of Cases	Death		Breast Cancer Death		Other Causes Death	
			N	%	N	%	N	%
Stage I	Control	120	27	22.50%	20	16.67%	7	5.83%
	Study	135	14	10.37%	10	7.41%	4	2.96%
	Refused	28	5	17.86%	2	7.14%	3	10.71%
Stage II	Control	94	52	55.32%	49	52.13%	3	3.19%
	Study	55	22	40.00%	22	40.00%	0	0.00%
	Refused	21	4	19.05%	4	19.05%	0	0.00%

N = no. of cases, % = percentage

40.00% in the study group. Thus the difference between two groups is not much. Here we can conclude that women will not benefit from screening if the breast cancer is detected in a later stage.

The tenth-year normalized proportion of cancer by detection stage is given in Table 6.12. Comparing Table 6.12 to Table 6.10 in the same way as before gives the number of women died in the second five years after detection. Again because of only few observations beyond stage II detection, the comparison is only conducted for stage I and stage II detection. Table 6.13 contains the number of women died either from breast cancer or from other causes in the second five years after detection. The proportion is taken by using the number of cases as the denominator as in Table 6.11.

From Table 6.13, during the second five years after detection for stage I detection in the control group, there are 11 women or 9.17% died from breast cancer. While in the study group, 17 women or 12.59% died from breast cancer. However for stage II detection, 14 women or 14.89% died from breast cancer in the control group and 12 women or 21.82% died from cancer. Here we have more breast cancer deaths in the study group. One plausible explanation for this phenomenon is that there is some kind of death delay effect in the study group; that is, some women who survived the first five years might not have been really cured. Instead the treatment they received only prolonged their life without

Table 6.12: Breast Cancer by Detection Stage - Normalized for Year 10

Stage	Control		Study		Refused	
	N	%	N	%	N	%
Stage I	80	66.67%	96	81.36%	17	54.84%
Stage II	27	22.50%	17	14.41%	9	29.03%
Stage III	4	3.33%	1	0.85%	1	3.23%
Stage IV	0	0.00%	0	0.00%	0	0.00%
Clinical Stage I or II	9	7.50%	4	3.39%	4	12.90%
Unknown	0	0.00%	0	0.00%	0	0.00%
Total	120		118		31	

N = no. of cases, % = percentage

Table 6.13: Number of Death during the Second Five Years after Diagnosis

Stage and Group		No of Cases	Death		Breast Cancer Death		Other Causes Death	
			N	%	N	%	N	%
Stage I	Control	120	13	10.83%	11	9.17%	2	1.67%
	Study	135	25	18.52%	17	12.59%	8	5.93%
	Refused	28	6	21.43%	0	0.00%	6	21.43%
Stage II	Control	94	15	15.96%	14	14.89%	1	1.06%
	Study	55	16	29.09%	12	21.82%	4	7.27%
	Refused	21	8	38.10%	8	38.10%	0	0.00%

N = no. of cases, % percentage

actually curing them. Another plausible explanation is that these might be a recurrent cases, including women with a new tumor in the other breast.

Chapter 7

Lead Time Analysis

Suppose there is no screening, then breast cancer will be detected, say at age V. With screening and assuming it is beneficial, the breast cancer will be detected at an earlier age, say Y. We shall call V the actual detection time. Some researchers, for example Shapiro et al (1982) and Xu et al (1995), called the difference $V - Y$ the lead time. The lead time, if it exists, is a relevant factor in assessing the efficacy of screening. In comparing the survival times of cancer patients in the study and the control groups, the lead time should be removed first. Otherwise, the possible longer survival times for the study group may be just an artificial effect and can be attributed to the lead time. Note the the lead time so defined is not an observable quantity, since V does not exist for a woman whose breast cancer was screening detected. Several stochastic models have been proposed in the literature to analyze the lead time effect. See, for example Xu et al (1995) and Xu et al (1999).

Table 7.1: Average and Standard Deviation of Detection Age

Group	N	Average	Std Dev
Control	268	54.578	6.6065
Study	211	54.483	6.5533
Refused	71	53.915	6.0018

Std Dev = standard deviation

Our approach to studying the lead time is to compare the distributions of the age at detection of breast cancer among the three groups.

First, ANOVA is conducted to test whether the means of detection age among groups are significantly different. The average and standard deviation of detection age for each group are given in Table 7.1. ANOVA shows no significant difference in the means (the F statistic is 0.29 with 2 degree of freedom which produces a p-value of 0.7456).

Next, we examine the binned data (according to age groups) of the age distribution at detection as given in Table 7.2. The test of equality of three multinomial distributions also showed no significant difference among the three groups. The chi-square statistic from this test is 7.2993 with 8 degree of freedom which gives a p-value of 0.5047.

Our final analysis is to use a "shift model" to investigate the lead time. Let Y be the age of detection of a cancer patient in the study group and X be the age of detection of a cancer patient in the control group. Set the expected lead time

Table 7.2: Table of Detection Age Group

	Control		Study		Refused	
Detection Age Group	N	%	N	%	N	%
Detection age ≤ 45	26	9.70%	21	9.95%	5	7.04%
$46 \leq$ detection age ≤ 50	60	22.39%	44	20.85%	20	28.17%
$51 \leq$ detection age ≤ 55	56	20.90%	50	23.70%	14	19.72%
$56 \leq$ detection age ≤ 60	62	23.13%	57	27.01%	22	30.99%
Detection age ≥ 61	64	23.88%	39	18.48%	10	14.08%
Total	268		211		71	

N = no of cases, % = percentage

$E(X - Y) = \theta$. We suppose that $X = Y + \theta$. Then the distribution function of X is $F(x - \theta)$ where F is the distribution function of Y. If the lead time indeed exists, then $\theta > 0$.

The empirical distribution functions for detection age (measured in years) for the control group, the study group and the refused group are given in Figure 7.1. The three empirical distribution functions are very close to one another. A estimate of lead time θ would be practically zero.

In summary, in all the statistical analysis that we conducted, we found no compelling evidence of the existence of significant lead time.

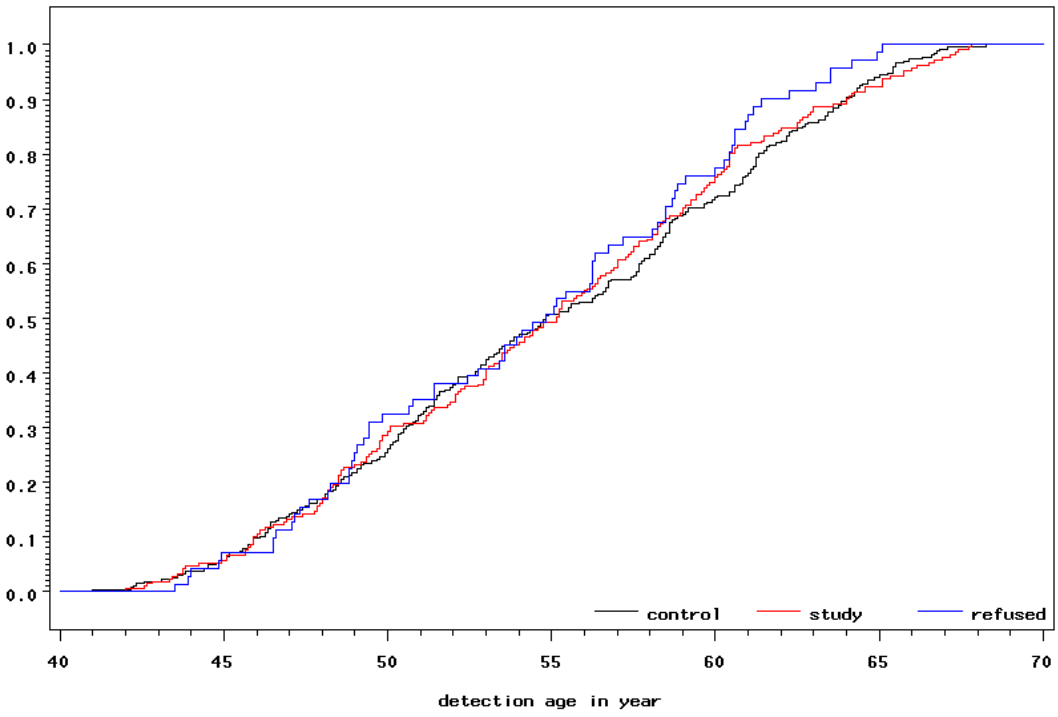


Figure 7.1: Empirical Distribution Function of Detection Age

Chapter 8

The Markov Chain

For future work, we would like to propose a Markov chain model to compare the study group and the control group. In our opinion, an appropriate Markov chain would take into account a multitude of transition patterns of the breast patients over the course of the disease development. We had made such an attempt but quickly realized that it was impossible to use Markov chain to analyze the the HIP data, because the HIP data did not record the dates of patients' remission and recurrences. Such information is needed for Markov chain analysis.

Nevertheless, we will present in this chapter such a Markov chain model and hopefully there will be more suitable data for analysis in the future. The Markov chain model that we propose is a four-state Markov chain. We construct one four-state Markov chain to model the transitions of each individual in the study group. Another four-state Markov chain will be used to model transitions of each individual in the control group and refused group.

For the study group, the four states are labelled by $\{0, 1, 2, 3\}$, where 0 is the disease free (or healthy) state, 1 is the pre-clinical disease state where an individual unknowingly has disease which can be detected by screening, 2 is the death state due to causes unrelated to breast cancer, and 3 is the death state caused by breast cancer.

For the control group, the four states are labelled by $\{0, 1, 2, 3\}$, where 0 is the disease free state, 1 is the clinical disease state where the cancer is detected without screening, 2 is the death state due to causes unrelated to breast cancer and 3 is the death state caused by breast cancer.

Individuals in the refused group follow a similar Markov chain as those in the control group.

It is obvious that states 2 and 3 are the absorbing states, while an individual can make transition from state $0 \to 1 \to 0 \to 1$ and so on before finally enters state 2 or 3.

A simple Markov chain may be constructed as follows. Assume that an individual with breast cancer is always diagnosed before death (an individual cannot go from state 0 to state 3 directly). Furthermore, for simplicity assume that an individual will not die from causes other than breast cancer (an individual cannot go from state 1 to state 2). The diagrams of the states are as follow

$$2 \leftarrow 0 \leftrightarrow 1 \to 3$$

We can generalized this model to include breast cancer cases that are not diagnosed before death and also the possibility that individuals can die from other causes.

If screening is advantageous, then the time to pre-clinical detection will be shorter than time to clinical detection. That is the sojourn time from state $0 \rightarrow 1$ in the study group will be shorter than the sojourn time from state $0 \rightarrow 1$ in the control group. Consequently the curing rate, which defined as the transition rate from state $1 \rightarrow 0$ in both the study group and the control group, will be bigger in the study group compare to the control group.

Chapter 9

Conclusion

Assessing the efficacy of screening for the breast cancer is indeed not an easy task. There are problems in every level of investigation, starting from the data collection up to the data analysis.

The most common study design used in this type of investigation is a randomized control trial. One argument that is often raised is how random is random. In the HIP project, at the beginning of the study there are twice as many women excluded from the project in the study group compared to the control group. This fact often raises question about bias in including/excluding women from the project. Aside from the bias in including/excluding women issue, one major problem in HIP is the compliance of the women in the project. In the data analysis, the major problem is selecting the start point and end point for the purpose of comparing the study group and the control group and treatment of the death due to other causes.

The HIP project does have its own advantages, because it is possibly one of the largest randomized controlled trial that was conducted when breast cancer screening is not as popular as it is now. One problem in recent breast cancer screening studies is that women in the control group are more likely conducting their own independent breast examination. In some studies, they are even told to conduct self-examination.

The key to better breast cancer survival seems to lie in early detection of the cancer. During the screening period, there are 0.877% breast cancer cases detected in the control group as oppose to 1.047% cases in the study group, a 0.17% difference. This difference in proportions is not statistically significant. Further investigation shows that the study group had a significantly higher stage I detection compare to the control group (50.00% in the control group as opposed to 67.77% in the study group). Analysis also shows that the study group has a better survival probability and also a lower death from breast cancer proportion compared to the control group. This proves that screening results in a higher early stage detection proportion in study group which in turn leads to better survival and less breast cancer death.

Because of these reasons, we conclude that breast cancer screening is beneficial for women. However, we would also remark that breast cancer screening in the HIP study consists of both mammography and examination by surgeons. Our conclusion does not imply that mammography alone is also advantageous.

BIBLIOGRAPHY

[1] Shapiro, S., Venet, W., Strax, P., Venet, L., Roeser, R. Ten- to Fourteen-Year Effect of Screening on Breast Cancer Mortality. Journal of National Cancer Institute 1982; 69:349-355.

[2] Shapiro, S. Periodic Screening for Breast Cancer: The HIP Randomized Controlled Trial. Journal of National Cancer Institute Monograph 1997; 22:27-30.

[3] Aron, J. L., Prorok, P. C. An Analysis of the Mortality Effect in a Breast Cancer Screening Study. International Journal of Epidemiology 1986; 15:36-43.

[4] Blanks, R. G., Moss, S. M., McGahan, C. E., Quinn, M. J., Babb, P. J. Effect of NHS Breast Screening Programme on Mortality from Breast Cancer in England and Wales, 1990-8: Comparison of Observed with Predicted Mortality. British Medical Journal 2000; 321:665-669.

[5] Gotzsche, P. C., Olsen, O. Is Screening for Breast Cancer with Mammography Justifiable? Lancet 2000; 355:129-134.

[6] Chu, K. C., Smart, C. R., Tarone, R. E. Analysis of Breast Cancer Mortality and Stage Distribution by Age for the Health Insurance Plan Clinical Trial. Journal of National Cancer Institute 1988; 80:1125-1132.

[7] Shapiro, S. Determining the Efficacy of Breast Cancer Screening. Cancer 1989; 63:1873-1880.

[8] Xu, J-L., Prorok, P. C. Non-Parametric Estimation of the Post-Lead-Time Survival Distribution of Screen-Detected Cancer Cases. Statistics in Medicine 1995; 14:2715-2725.

[9] Xu, J-L., Fagerstrom, R. M., Prorok, P. C. Estimation of Post-Lead-Time Survival Under Dependence between Lead-Time and Post-Lead-Time Survival. Statististics in Medicine 1999; 18:155-162.

[10] Miller, R. G., Survival Analysis. New York, Wiley, 1981.

[11] Yang, G. L., Chang, M. N. A Stochastic Model for Analyzing Prevalence Surveys of Hepatitis A Antibody. Mathematical Biosciences 1990; 98:157-169.

[12] Zelen, M., Optimal Scheduling for the Early Detection of Disease. Biometrika 1993; 80:279-293.

[13] Ross, S., Stochastic Processes. New York, Wiley, 1996.