POINT ESTIMATES AND BOOTSTRAP INTERVALS OF LEAD TIME IN THE NATIONAL BREAST SCREENING STUDY

By

Helen Xun Xie

A thesis submitted in conformity with the requirements for the degree of Master of Science
Graduate Department of Community Health
University of Toronto

© Copyright by Helen Xun Xie, 1997
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.

L’auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

L’auteur conserve la propriété du droit d’auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

0-612-28756-4
Point Estimates and Bootstrap Intervals of Lead Time in the National Breast Screening Study

Degree: Master of Science
Year of Convocation: 1997
Helen Xun Xie
Department of Community Health
University of Toronto

Abstract

The Walter and Day lead time estimation methodology is applied to the National Breast Screening study (NBSS) to estimate the mean lead time; the underlying incidence rate of breast cancer in the population; and the false negative rate of screening test. This thesis is the first attempt to systematically evaluate the properties of estimated screening parameters, construct their confidence intervals and measure the associated biases of estimates using the bootstrap method. This thesis suggested that Walter and Day’s lead time estimation methodology can be used as a tool in the evaluation of screening. Confidence intervals generated for the lead time and the incidence rate by the bootstrap method are relatively narrow, and therefore precise compared to wider confidence interval around the false negative rate. Future research is necessary as questions relating to the precision of the false negative rate estimated by Walter and Day’s method remain.
Acknowledgements

This thesis used data and resources from the Canadian National Breast Screening Study (NBSS) which is supported by the National Cancer Institute of Canada (NCIC).

I would like to thank a number of people who were helpful throughout the thesis process. Many thanks to Dr. Teresa To, supervisor, for overall guidance, support and insight and for statistical and methodological advice and encouragement. I would also like to thank my committee members, Drs. Paul Corey and Robert Tibshirani, and my examiners, Drs. Anthony Miller and Gerrit DeBoer, for their helpful comments and suggestions. Finally, I would like to thank Mr. Andrew White, from the NCIC for extensive computer help, including teaching me FORTRAN and Mrs. Catheline Andrew for answering my queries concerning the NBSS.

I am also grateful to friends, family, especially my husband, Zongde Lin, and my parents for emotion support, understanding, encouragement, and love throughout my studies.
Contents

Abstract ii

Acknowledgements iii

List of Tables vii

List of Figures viii

1 Introduction 1

2 Literature Review 5

2.1 Schematic illustration of the disease progression with screening 5

2.2 Biases in a screening program 7

2.3 Sensitivity and Specificity 10

2.4 Methods of estimating parameters of preclinical disease with a screening program 12

2.4.1 Estimates based on the constant duration of preclinical disease 12

2.4.2 Estimates of some characteristics of the disease natural history by using a probabilistic model 17

2.4.3 The Stochastic Model of screening for progressive disease 26

2.4.4 Other Approaches 28

2.5 Summary 30
3 Materials and Methods

3.1 Data source - the National Breast Screening Study (NBSS)  
3.2 Study objectives  
3.3 Walter and Day's methods used to estimate the incidence rate, the false negative rate, and the lead time  
3.4 Bootstrap process  
3.5 Bootstrap confidence intervals  
3.5.1 Confidence intervals based on bootstrap percentile  
3.5.2 The bias-corrected and accelerated (BC_a) intervals  
3.6 Bootstrap estimate of bias  
3.7 Grid search method  
3.8 Application

4 Results

4.1 Point estimates of incidence rate, false negative rate, and lead time based on the NBSS data

4.2 Interval estimates of incidence rate, false negative rate, and lead time based on the NBSS data by the bootstrap method

4.3 Bias estimates of incidence rate, false negative rate, and lead time based on the NBSS data

5 Discussion

5.1 Discussion of the methods

5.1.1 Advantages and limitations of point estimates of association in this thesis

5.1.2 Advantages and limitations of interval and bias estimates of association in this thesis

5.2 Direction for future research

Bibliography
Appendices

A.1 Key terms ......................................................... 89
A.2 Expressions for the post-screening incidence: two screens take place at
times $t_1$ and $t_2$ .................................................. 90
A.3 Expressions for the prevalence at screening: two screens take place at
times $t_1$ and $t_2$ .................................................. 92
A.4 The detail procedures for results presented in Tables 4.1 to 4.8 .... 95
List of Tables

2.1 Summary of lead time estimates of previous studies ... ... ... ... ... 33

4.1 Observed and Expected Number of Cancers (MA 40-49) ... ... ... 59
4.2 Observed and Expected Number of Cancers (PE 40-49) ... ... ... 60
4.3 Observed and Expected Number of Cancers (MA 50-59) ... ... ... 61
4.4 Observed and Expected Number of Cancers (PE 50-59) ... ... ... 62
4.5 Observed and Expected Number of Cancers (MA all ages combined) 63
4.6 Observed and Expected Number of Cancers (PE all ages combined) 64
4.7 Observed and Expected Number of Cancers (All 40-49 combined) 65
4.8 Observed and Expected Number of Cancers (All 50-59 combined) 66
4.9 Parameter Estimates Based on the Exponential Distribution ... ... 67
4.10 Parameter Estimates Based on the Weibull Distribution ... ... ... 67
List of Figures

2.1 Progression of a disease .................................... 6

3.1 Schematic Illustration of Bootstrap Process ................. 40

4.1 Histogram of 1,000 bootstrap replications of \( \hat{I} \), incidence rates, for the MA 40-49 ........................................... 68

4.2 Histogram of 1,000 random numbers from the normal distribution \( N(0.0023, 2.15E-08) \) for the MA 40-49 ......................... 68

4.3 Histogram of 1,000 bootstrap replications of \( \hat{\theta} \), false negative rate, for the MA 40-49 ........................................... 69

4.4 Histogram of 1,000 random numbers from the normal distribution \( N(0.197, 0.0138) \) for the MA 40-49 ......................... 69

4.5 Histogram of 1,000 bootstrap replications of \( 1/\hat{\lambda} \), the lead time, for the MA 40-49 ........................................... 70

4.6 Histogram of 1,000 random numbers from the normal distribution \( N(0.438, 0.0074) \) for the MA 40-49 ......................... 70

4.7 Histogram of 1,000 bootstrap replications of \( \hat{I} \), incidence rates, for the MA 50-59 ........................................... 71

4.8 Histogram of 1,000 random numbers from the normal distribution \( N(0.0029, 3.88E-08) \) for the MA 50-59 ......................... 71

4.9 Histogram of 1,000 bootstrap replications of \( \hat{\theta} \), false negative rate, for the MA 50-59 ........................................... 72
4.10 Histogram of 1,000 random numbers from the normal distribution
N(0.277, 0.0070) for the MA 50-59 ........................................ 72
4.11 Histogram of 1,000 bootstrap replications of 1/%, the lead time, for the
MA 50-59 ............................................................................. 73
4.12 Histogram of 1,000 random numbers from the normal distribution
N(0.2748, 0.0022) for the MA 50-59 ........................................ 73
4.13 Histogram of 1,000 bootstrap replications of ũ, incidence rates, for the
PE 50-59 ............................................................................. 74
4.14 Histogram of 1,000 random numbers from the normal distribution
N(0.0027, 3.21E-08) for the PE 50-59 ........................................ 74
4.15 Histogram of 1,000 bootstrap replications of Ô, false negative rate, for
the PE 50-59 ........................................................................ 75
4.16 Histogram of 1,000 random numbers from the normal distribution
N(0.3300, 0.0208) for the PE 50-59 ........................................ 75
4.17 Histogram of 1,000 bootstrap replications of 1/%, the lead time, for the
PE 50-59 ............................................................................. 76
4.18 Histogram of 1,000 random numbers from the normal distribution
N(0.6614, 0.344) for the PE 50-59 ........................................ 76
Chapter 1

Introduction

Recent trends indicate that breast cancer is responsible for more new cases of cancer among women than any other cancer. It is the number one cause of death for Canadian women aged 35 - 64 years [1] and the second cause of cancer deaths among American women (surpassed only by lung cancer) [2]. Among caucasian women under the age of 50, there appears to be a slight but consistent increase in the percentage of cases with localized disease at diagnosis: the trends in the stage distribution among caucasian women over the age of 50 at diagnosis are much clearer. There has been a large increase in the percentage of cases with localized disease at diagnosis. Breast cancer is the most common cancer in women and it is the leading "killer cancer" in women of any age. It is a devastating disease, and its incidence increases with age. Breast cancer becomes one of the most serious problems to North American women today. It is, therefore, important to have earlier detection and earlier treatment of the disease as breast cancer is curable and appropriate therapies prolong life [2].

Early detection of some chronic diseases is achieved through the use of special diagnostic procedures along with the establishment of public health programs. For instance, tuberculosis can be detected by using chest X-rays; women with cancer of the uterine cervix can be detected by using Pap smears; other programs designed to test for glaucoma and diabetes are widely in use. An especially interesting program
for early detection of breast cancer using soft tissue X-rays, mammography, has been conducted by the National Breast Screening Study for many years. In short, the objective of all such screening programs is to have the most effective weapon, through early detection and early treatment, to discover and cure those among the apparently well who are in fact suffering from disease. The disease involved should in a practical way stand a reasonable chance of being cured or having a better prognosis if detected early: whereas, if it is not diagnosed until the patient comes to the physician with clear-cut symptoms, it may be incurable. Delay in diagnosis may eliminate the chance of cure. As a result, early detection and early application of appropriate therapy for breast cancer has lead to a lower mortality rate, particularly in women over 50 years of age [3]. Patients do live longer and live well with their disease [2].

The potential benefit of a screening program is related to the lead time gained by early diagnosis. Lead time is the time by which the diagnosis is advanced by using a special diagnostic procedure. Lead time can be used for at least two purposes in the evaluation of a screening program. First, it is a measure of how much earlier than usual a disease is discovred. Second, it can be applied as a correction factor in the comparison of a screened population with a control population when examining survival experience or case fatality rate [4]. Lead time can be estimated by both direct and indirect ways. The direct way is to observe the length of time until the individual enters the clinical state, if she is found to have preclinical disease by screening. Obviously, this procedure raises serious ethical questions and must be discarded. The indirect way to estimate lead time requires that an individual having the disease be given therapy after detection [5]. Lead time is not directly observable once screening is introduced. How to use observable information to estimate unobservable parameters becomes one of the purposes of this thesis, however, other characteristics of the screening program should be evaluated as well. An effective screening tool must have the ability to detect cancer at an earlier stage than would otherwise occur. The false negative rate, another component of the screening program, is defined as the ability of the screening tool to detect cancers in the preclinical stage. A negative report from
a screening examination, before the time a case reached the clinical level, can result in delayed detection of the disease. Minimizing the false negative rate will maximize sensitivity which will ensure the detection of a maximum number of cases, and therefore, appropriate biopsies or treatment can be prescribed in a timely manner.

It is a widely accepted belief that most cancers are better treated earlier than later. Because early detection interrupts the natural course of a disease process, it is not possible to answer the question, that is, if earlier detection is associated with better patient prognosis, empirically on a patient-by-patient basis. Instead it becomes necessary to compare the statistical behavior of the disease's temporal characteristics in screened and comparable unscreened populations. In order to make quantitative predictions about a screening program's effect, it is necessary to develop a mathematical model which relates the statistical behavior of a population's disease natural-history characteristics to the characteristics of the screening program.

The Walter and Day lead time estimation methodology provides, simultaneously, the probability distribution of the duration of the detectable preclinical phase; the incidence rate of disease in the population; and the screening false negative rate. This methodology was applied to the Health Insurance Plan of Greater New York (HIP) Study of Breast Cancer. Unlike Walter and Day's method, most literature derive lead time estimation based on different assumptions and applied to the HIP data, but do not simultaneously consider all biases such as sample selection bias, lead time bias, length-biased sampling, and false negative rate.

In this thesis, Walter and Day's methodology is applied to the data abstracted from the Canadian National Breast Screening Study (NBSS) to obtain point estimates of the lead time, the incidence rate (I), and the false negative rate (θ). The properties and characteristics of these estimates have not been previously tested. This thesis represents an effort to apply Walter and Day's method to a large database (NBSS data), and it is possibly the first attempt to systematically evaluate the properties of the estimated parameters by constructing confidence intervals for those associated
parameters using the bootstrap method.

The objectives of this thesis are (1) to obtain simultaneously the point estimates of the mean duration of the detectable preclinical phase, the false negative rate ($\theta$) of the screening test, and the underlying incidence rate ($\lambda$) in the NBSS population using Walter and Day's methodology; (2) to estimate the lead time for different allocations and age groups; (3) to construct the confidence intervals for the lead time, $\theta$, and $\lambda$ by using the bootstrap percentile method; (4) to assess the reliability of the estimates of the parameters $\lambda$, $\theta$, and $\lambda$ and to measure the associated biases.

Chapter 2 describes the literature review. This chapter starts with a brief description of the disease progression of a screened individual. Some biases in the screening program are discussed. Different approaches to estimate parameters in the screening tests are also reviewed. Data source and details of the methodology used in this thesis are described in Chapter 3. Chapter 4 explains the results based on the NBSS data. Finally, a general discussion of the results is given in Chapter 5. The problems and limitations of the thesis are discussed, as well suggestions for future directions in research.
Chapter 2

Literature Review

2.1 Schematic illustration of the disease progression with screening

Figure 2.1 shows a simplified schematic representation for the disease progression of an individual with screening. A person who initially lives a certain portion of his/her life, free of the type of disease (cancer), may move into a certain point, \( T_1 \), at which time a primary disease begins to develop, referred to as the biological onset of the disease. At this time, the subject leaves the disease free state and enters an asymptomatic disease stage which is usually not recognized by the patient and not even detectable by conventional medical techniques (e.g., screening). With the passage of time, the disease progresses through various phases. When the disease moves into point \( T_2 \), referred to as the point at which the disease becomes detectable by screening, if screening is used, the disease can be detected; if not, the disease will progress to symptomatic stage at time \( T_3 \), the usual diagnosis time when the subject expresses symptoms and seeks medical help. The onset of breast cancer is initiated by a change in a single cell at time \( T_1 \), and that after a certain duration it reaches a size that is presented to medical attention and recognized clinically as a palpable mass at time \( T_3 \). Before such recognition, there is an interval \( (T_2, T_3) \) in which the lesion
may be detected in a screening program. This early stage of the disease, recognizable only in some special screening examinations, is referred to as preclinical disease. The interval between $T_2$ and $T_3$ constitutes the detectable preclinical phase (DPCP) of the disease, denoted by $X=T_3-T_2$. If screening takes place at time $t$ between $T_2$ and $T_3$, and the disease is detected, then the intervals $L=T_3-t$ and $D=t-T_2$ represent the lead time and the detection delay time respectively. $X$ is the sum of the lead time and the delay time.

\[ X = T_3 - T_2 \]

Figure 2.1: Progression of a disease
2.2 Biases in a screening program

Careful design of a screening program is necessary because of specific biases that may occur, as described below.

Lead time is the period of time between the detection of a medical condition by screening and when it ordinarily would have been diagnosed because an individual experienced symptoms and sought medical care. Lead time is unobservable for a particular case detected by screening. The distribution of the lead time for all such cases depends on the distribution of the detectable preclinical phase (DPCP). Since clinical breast cancer is known to vary widely from individual to individual in the rate of progression, the distribution of the preclinical duration therefore probably has a wide variation. As a result, lead time also varies from individual to individual. The amount of lead time for a given disease is dependent on both the biologic rate of progression of the disease and on the ability of the screening test to detect early disease. If DPCP is long, then the maximum attainable lead time is correspondingly long. If DPCP is short, however, the potential gains from screening are limited, and screening must then take place more frequently in order to increase the probability that the preclinical disease is detected before it becomes clinically apparent. When lead time is very short, the treatment of medical conditions detected on screening is likely to be no more effective than treatment after symptoms appear. On the other hand, when lead time is long, the treatment of medical conditions detected by screening can be very effective.

Since screening allows a disease to be detected early, in theory those cases who are diagnosed by screening for a deadly disease will, on the average live longer from the time of diagnosis than people who are diagnosed after they expressed symptoms. Even if there is no effective treatment, it could appear that screening improves survival. when in reality, it is the 'disease time' that is extended as opposed to 'survival time'. Lead time could result in a bias, which is referred to as the lead time bias. in studies using survival as a measure of the efficacy of early treatment.
An appropriate method to adjust for the lead time bias is to study both a screened group of subjects and a comparable control group and to compare their age-specific mortality rates, rather than survival rates from the time of diagnosis. Survival studies for cancer patients who have been detected by screening programs will be distorted by the lead time effect if survival times are measured from the time of diagnosis.

The other major problem in the evaluation of screening, is that the cases of disease which are detected by the screen are not representative of all cases in the population. One reason for this is that certain selection factors may produce a biased sample of those eligible by affecting an individual's acceptance of screening. A screened population is a group with abnormally high susceptibility to breast cancer, perhaps reflecting the inclusion of some individuals with breast cysts, pain, or lumps, who accepted the screening because of symptoms, or who knew they had risk factors for the disease. Such factors, if related to the disease etiology, will confound the disease experience in the screened persons to be different from the group of unscreened persons. If a study compares disease outcomes among volunteers for a screening program with outcomes in a group of subjects who were not volunteers, better results observed in the volunteers may be due to other differences between the two groups of patients rather than to treatment effect. Fortunately, this bias, referred to as the sample selection bias, can be eliminated by randomization of study or control groups.

Another reason that prevalent screened cases are non-representative is the phenomenon of the length-biased sampling. Length-biased sampling occurs because the proportion of slow-growing lesions diagnosed during screening programs is greater than the proportion of those diagnosed during usual medical care. The result of including a greater number of slow-growing cancers can make it appear that screening and early treatment are more effective. Most types of cancers demonstrate a wide range of growth patterns. Some cancers grow slowly, some very fast. Screening tests are likely to find mostly slow-growing tumors, because fast-growing ones will more likely have already expressed symptoms that lead to diagnosis in the interval between screening
examinations. As screening is intermittent, rarely more frequently than every year, screening, therefore, tends to find tumors with better prognosis, whereas regular care finds those with worse prognosis, even if the prognosis of each is unchanged by subsequent treatment. As a result, the mortality rates of cancers found on screening may appear better than those not found on the screening, but it is not because of the screening itself. Screening will be overweighted with cases having long preclinical duration. cases that have a biologically slower progression and have a somewhat better prognosis than persons with more aggressive disease.

Bias due to length/time can be avoided by relying on randomized controlled trials that count all the outcomes in the group, regardless of the method of diagnosis or degree of participation. Groups of patients that are randomly allocated will have comparable number of slow and fast-growing tumors.

Most of previous analyses assumed that all cases found at screening would eventually have presented clinically, so that prevalent and incident cases can be considered on an equal footing. In fact, in some situations, the lesions detected may well not progress to invasive disease. Under this circumstance, the analyses should be confined to the fewer incidence cases. It is a kind of overdiagnosis, one of the screening biases, which may affect the lead time estimation. There is likely to be substantially more overdiagnosis from mammography than physical examination screening.

All such biases, i.e., lead time bias, length-biased sampling, selection bias, and overdiagnosis must be taken into account when comparing clinical outcomes in cases found by screening with other clinical incident cases. For example, when computing survival of screen-detected cases, one must allow for the lead time which may produce an earlier diagnosis, but not necessarily a later death. Similarly, when estimating lead time, one must take the length-biased sampling, selection bias and overdiagnosis into account to avoid erroneous conclusions [7].
2.3 Sensitivity and Specificity

Sensitivity is defined as the proportion of people with disease who have a positive test for the disease. A sensitive test will rarely miss people with the disease. Specificity is the proportion of people without the disease who have a negative test. A specific test will rarely misclassify people without the disease as cases. Both sensitivity and specificity are aspects of the validity of a test. An effective screen must have the ability to detect cancers at an earlier stage than would occur otherwise. Sensitivity measures the ability of a screening test to correctly label as positive those who do have the disease in the DPCP. Specificity measures the ability of a screening test to correctly label as negative those who do not have the disease. A test is more valid if both the sensitivity and specificity of the test are maximized. Sensitive tests are useful when the probability of disease is relatively low and the purpose of the test is to discover the disease.

The very nature of searching for disease in asymptomatic people means the prevalence of a particular disease is usually very low, even among high risk groups selected because of age, sex, and other characteristics. A good screening test must, therefore, have a high sensitivity in order not to miss the few cases of disease that are present, and a high specificity, to reduce the number of people with false positive results who require further tests. Sometimes, there is a trade-off between sensitivity and specificity, with the need to balance the extremes of maximizing sensitivity which minimizes specificity, or maximizing specificity which minimizes sensitivity. Maximizing sensitivity will ensure that a maximum number of cases is detected which will minimize the number of false negatives. A negative report from a screening examination before the time a case reached the clinical level can result in delayed detection of the disease. Minimizing the false negative rate, for breast cancer, will therefore, ensure necessary biopsies are offered to women detected as cancer patients. Maximizing specificity will minimize the number of false positives for breast screening, therefore, minimizing the number of women who are given unnecessary biopsies.
Medical screening procedures for early detection of disease are often used in large populations before their clinical effectiveness has been evaluated. Knowledge of the true disease status for all screened individuals would allow estimation of the false negative and false positive rates for each mode of detection and for the program as a whole. The method of using data from a screening program to estimate the probabilities that the screen will miss a "disease" person or incorrectly identify a healthy one has been discussed in [6]. Misclassification rates are needed to determine whether screening is operating effectively as a classifier. The first time that screening is carried out - the prevalence screen, cases of the medical condition will be present for varying lengths of time. During the second round of screening, most cases found will have had their onset between the first and the second screening. Therefore, second and subsequent screenings are called incidence screens. When a group of people are periodically rescreened, the number of cases of disease present in the group drops after the prevalence screen.

In practice, all individuals identified as positive on screening are followed and separated into true positives and false positives; individuals screened negative are not followed but return to the population at risk. Further, an individual screened negative who is followed and subsequently found to be positive may represent either a false negative on screening or a newly developed case.

Complete evaluation of all the screening programs require estimation of the mean lead time by which diagnosis is advanced. Mean lead time has been used to evaluate screening programs for chronic disease by many authors. Proposed estimators of mean lead time depend on the false negative rates of the screening mode; therefore, estimation of these rates is essential to evaluate a screening program.

Unfortunately, information on negative tests, whether true negative or false negative, tends to be much less complete in the medical literature.
2.4 Methods of estimating parameters of preclinical disease with a screening program

The stage that is recognizable only through special screening efforts is referred to here as preclinical disease. Since clinical disease is known to vary widely from patient to patient, it is reasonable to conclude that preclinical disease has a highly variable rate of progress as well and, that cases detected through screening vary in the duration of their preclinical state. Although the duration of the preclinical phase is unobservable for both prevalence and incidence cases, knowledge of the distribution of cases by preclinical duration, its relationship to personal characteristics, and the screen sensitivity are very important for determining the potential benefits and optimal screening strategy. Thus, it is necessary to estimate the distribution of duration of preclinical disease (X). The data provide a basis for estimating the prevalence, the incidence, the mean duration of preclinical disease, and the average lead time attribution to the screening program.

A useful method to evaluate the fit of observed data with various assumed possibilities for the distribution of X is given by Walter and Day [7]: once the distribution having the best fit has been determined, the implied distribution of lead time can be calculated. Several other statistical models have been applied in the current study to derive these estimates, which are described in the following sections.

2.4.1 Estimates based on the constant duration of preclinical disease

One approach to estimate the parameters of preclinical disease, suggested by Shapiro, is based on the assumption that all cases have the same total duration of preclinical disease [8,9,10]. Under this assumption, the average duration to the date of screening and mean lead time are both one-half the total duration which is estimated by the ratio of prevalence at the first screen to the estimated incidence in the screened group.
Shapiro used four types of mean to determine the incidence of preclinical breast cancer applied to the Health Insurance Plan of Greater New York (HIP) data. The four types of mean were: 1) the incidence of clinical breast cancer in some general population, i.e., incidence from prior study; 2) the incidence of the control group of the study; 3) the incidence of preclinical disease estimated directly by following up disease-free women for a time period and then screening; and 4) the incidence of clinical breast cancer determined in unscreened women in the study population. These four estimates were compared, and they implied that each of the estimates from unscreened women (control group and unscreened group) are individually comparable with the ranges of estimates predicted from the prior study. The rate for the unscreened women in the study group, however, is significantly lower than either the control group rate or the rate for the screened women. This implied that the screened and unscreened subgroups of the study group are not similar subgroups, but rather the screened women belong to a higher breast cancer risk category.

The incidence of the total study population if nobody has been screened can be expressed as follows:

\[ I_t = \frac{I_s N_s + I_{ns} N_{ns}}{N_t} \]

Where \( I_s, I_{ns} \) and \( I_t \) are incidence rates for screened, unscreened women in the study group, and the total study group, respectively. \( N_s, N_{ns}, \) and \( N_t \) are the corresponding numbers of person-years for screened, unscreened women in the study group, and the total study group, respectively. It is assumed that the underlying incidence rate for the total study group is the same as the rate for the total control group. Since the incidence rate for women who refused to participate in the screening program is the same rate as unscreened women in study group, this can be obtained by following them up for a time period, expressed in below:
Where $I_c$ is the incidence rate for the control group, and $N_c$ is the number of person-years for the control group.

Therefore, for a one-time screening program, the mean lead time is expressed as:

$$L = \frac{1}{2} \frac{P}{I_s}.$$  

Where $P$ is the prevalence rate for screened women.

Assuming that preclinical disease normally varies in the total population, the prevalence cases found on screening have an average duration-to-date longer than half the average total duration of a group of incidence cases. This is because short duration cases of preclinical disease originating several years before, will already have become diagnosed clinical disease and, therefore, will not present as preclinical disease at survey. Long duration cases originating in these same years, however, will still be preclinical and will be found. In general, when prevalence cases are compared with incidence cases, the prevalence will be relatively over represented in long duration cases and under represented in short duration cases. The average duration-to-date will be half the expected average total duration of these screened cases and therefore greater than half the expected mean total duration of all new cases, determined by $P/I$.

If screening is done more frequently, prevalence, incidence, and the lead time can be expressed as follows: Immediately following a screening, prevalence is zero and builds up gradually as cases newly detectable by screening accumulate. After an interval $t$ following screening, the prevalence is expressed as:
\[ P_t = \sum_{x=0}^{t} I_x \times x + t \sum_{x=t+1}^{\infty} I_x. \]

Where \( x \) is the duration, during which an individual case is diagnosable by screening but not diagnosed under usual practice. \( I_x \) is the incidence of cases of duration \( x \). Prevalence consists of two parts. One is short duration cases (duration \( \leq t \)), and the other is long duration cases (duration \( > t \)).

Incidence of cases diagnosed by usual procedures is zero immediately after screening and remain zero for a period \( \tau \), where \( \tau \) is the duration of the shortest duration case. After an interval \( t \), cases of all duration \( x \) will be diagnosed. The total incidence in interval \( t \) following screening is expressed as:

\[ I_t = \sum_{x=0}^{t} I_x(t - x). \]

The mean duration-to-date of a prevalence group diagnosed at screening with screening carried out at interval \( t \) following prior screening is expressed as:

\[ \bar{x}_1 = \frac{1}{2} \frac{\sum_{x=0}^{t} I_x \times x \times x + t \times t \sum_{x=t+1}^{\infty} I_x}{\sum_{x=0}^{t} I_x \times x + t \sum_{x=t+1}^{\infty} I_x} = \frac{1}{2} \frac{\sum_{x=0}^{t} I_x x^2 + t^2 \sum_{x=t+1}^{\infty} I_x}{\sum_{x=0}^{t} I_x x + t \sum_{x=t+1}^{\infty} I_x}. \]

and for \( t=\infty \),

\[ \bar{x}_1 = \frac{1}{2} \frac{\sum_{x=0}^{\infty} I_x x^2}{P} = \frac{P}{2} + \frac{Var(X)}{2P}. \]

Where \( Var(X) \) is the variance of the duration of the preclinical disease.

The mean duration-to-diagnosis of an incidence group diagnosed by usual procedures in an interval \( t \) following screening is expressed as:
\[ x_2 = \frac{\sum_{x=0}^{t} I_x (t - x) \times x}{\sum_{x=0}^{t} I_x (t - x)} . \]

For the prevalence series considered alone the mean duration is expressed as:

\[ \bar{x} = \frac{\sum_{x=0}^{t} I_x x^2 + t \sum_{x=t+1}^{\infty} I_x x}{\sum_{x=0}^{t} I_x x + t \sum_{x=t+1}^{\infty} I_x} . \]

The average lead time for the prevalence series is expressed as:

\[ L = \bar{x} - \bar{x}_1 = \frac{\frac{1}{2} \sum_{x=0}^{t} I_x x^2 + t \sum_{x=t+1}^{\infty} I_x x - \frac{t^2}{2} \sum_{x=t+1}^{\infty} I_x}{\sum_{x=0}^{t} I_x x + t \sum_{x=t+1}^{\infty} I_x} . \]

The average lead time for the incidence series considered alone is zero.

An alternative approach to estimate the average lead time is somewhat different. It avoids the assumption that the duration of preclinical disease is a constant, but it assumes that new cases arise at a constant rate. This method compares the observed and expected dates of diagnosis of the screened cases.

In the former approach, since the randomization of study or control group is used, the overall selection bias factor for the women who accept the screen is taken into account, but it ignores the assessment of the sensitivity of the screening program, including a review of mammogram negatives on screening for women with breast cancer detected between screening examinations, that is, it ignores the possibility of false negative results. It also assumes the duration of preclinical disease is a constant, and therefore ignores the length-biased sampling issue. Thus both procedures give estimates of average lead time that involve minimizing assumptions, and more precision cannot be gained until information on the distribution of preclinical duration becomes better known. Improved estimates can be determined by considering the distribution of the total duration of preclinical disease, which will be discussed in the following sections.
2.4.2 Estimates of some characteristics of the disease natural history by using a probabilistic model

In a series of papers [11.12.13], a basic probabilistic formulation for the natural history of a progressive disease in a population was presented through the use of a simple "disease-state" model (figure 2.1). It represents the mathematical foundation of a quantitative model for evaluating the natural cancer screening effort. The basic formulation is then used to derive mathematical descriptions of epidemiology measures, such as age specific incidence rates, prevalence rates, mean preclinical duration time, and lead time. The approaches, shown in the following sections, are based on the models which incorporate changes in the population over time due to migration, death, and screening.

2.4.2.1 Natural history of disease in a stable population

In Albert's paper [11], the author characterized the natural history of a chronic progressive disease such as cancer, in terms of the distribution of $Y$ (a person's age at the time of entering the detectable disease state $S$), $X$ (the sojourn time in that disease state $S$, i.e., DPCP), and $A$ (a person's current age) over a population of individuals. They presented a basic probabilistic formulation for it, through the use of a simple disease-state model. Considering a population of individuals, each associated with $(Y,X,A)$ values, at an instant of time $t$, they defined a joint distribution for $(Y,X,A)$, which can be estimated using a three-way frequency table of counts and represented by a density function, $f_{YX_A}(y,x,a,t)$. If the population is stable, i.e., no deaths, or migration, then the density of $(Y,X,A)$ at time $t+\mu$ is related to the density at time $t$ by $f_{YX_A}(y,x,a,t+\mu) = f_{YX_A}(y,x,a-\mu,t)$, since $Y$ and $X$ are unchanged over time. Based on this assumption, some mathematical description of traditional epidemiologic measures, such as age specific incidence, prevalence, and mean duration of the disease state, etc. at an instant time $t$ were derived.

The incidence rate is expressed as:
\[ I_s(t) = \int_0^\infty I_s(a) f_A(a) \, da. \]

Where \( I_s(t) \), the overall incidence of S at the instant \( t \), is the instantaneous rate at which people in the population (all ages) enter S. \( I_s(a) \), the age specific incidence of S among those aged a, is the instantaneous rate at which members of population whose present age is a enter S. \( f_A(a) \) is the marginal densities of A. In summary, the overall incidence of S is the weighted average of the age specific incidence of S with respect to the age distribution that obtains in the population at that instant.

The prevalence rate is expressed as:

\[ P_s(t) = \int_0^\infty P_s(a) f_A(a) \, da. \]

Where \( P_s(t) \), the overall prevalence of S at the instant \( t \), is the proportion of those in the population (all ages) who are in S. \( P_s(a) \), the age specific prevalence of the disease state S among those aged a, is the proportion of individuals, among those aged a at time \( t \), who are then in the disease state S. \( f_A(a) \) is the same as above. In other words, the overall prevalence of S is the weighted average of the age specific prevalence of S with respect to the age distribution that prevails over the population at that instant.

The mean duration of S (sojourn time) is expressed as:

\[ E(X|Y < \infty) = \frac{\int_0^\infty P_s(a) \, da}{\int_0^\infty I_s(a) \, da}, \]

which can be interpreted as the ratio of prevalence to incidence for those diseases in which the population age distribution is flat, but not for diseases which characteristically attack the aged.
The above formula is different from the traditional formula

\[ E(X|Y < \infty) = \frac{P_s}{I_s} = \frac{\int_0^\infty P_s(a)f_A(a)da}{\int_0^\infty I_s(a)f_A(a)da}. \]

The latter is based on the hidden assumption that no migration takes place in the population. In fact, a "uniform" age distribution implies a negligible net rate of death and migration and, therefore, allows the "classic" result to be obtained.

In the absence of screening, a cancer patient will seek medical care at age \( Y + X \) at which age his/her symptom appears. If the patient were to be diagnosed by screening, and if the treatment were to begin without delay, treatment would begin at some age between \( Y \) and \( Y + X \). Hence the distribution of lead time \( L \) is

\[
F_L(t) \equiv P_r[L > t|\text{Subject in preclinical state } S] = P_r[Y + X - A > t|Y \leq A < X + Y] = \frac{\int_0^\infty \int_{t+u}^\infty f_{X,A-Y}(x,u)dxdu}{\int_0^\infty \int_{u}^\infty f_{X,A-Y}(x,u)dxdu}.
\]

In the same fashion, the distribution of delay time \( D \) is

\[
F_D(t) \equiv P_r[D > t|\text{Subject in preclinical state } S] = P_r[A - Y > t|0 < A - Y < X] = \frac{\int_t^\infty \int_u^\infty f_{X,A-Y}(x,u)dxdu}{\int_0^\infty \int_u^\infty f_{X,A-Y}(x,u)dxdu}.
\]

Therefore, the mean lead time and mean delay time are

\[ E(L|\text{Subject in preclinical state}) = \int_0^\infty F_L(u)du \]
\[ E(D|\text{Subject in preclinical state}) = \int_0^\infty F_D(u)du. \]

If \( X \) and \( Y \) are independent, then \( Y+X-A \) and \( A-Y \) have a uniform distribution over \( X \), then

\[ E(L|\text{Subject in preclinical state } S) = \frac{1}{2} \frac{EX^2}{EX} = \frac{1}{2} EX[1 + \frac{\sigma^2(X)}{E^2(X)}] \]

\[ E(D|\text{Subject in preclinical state } S) = \frac{1}{2} \frac{EX^2}{EX} = \frac{1}{2} EX[1 + \frac{\sigma^2(X)}{E^2(X)}]. \]

If \( X \) has an exponential distribution then \( \sigma^2(X) = E^2(X) \) and so

\[ E(L|\text{Subject in preclinical state } S) = E(D| \text{Subject in preclinical state } S) = EX. \]

The above derived expressions are based on assumptions that screenings are performed all at once, i.e., instantaneously, and are error free (no false negative errors). Note that with the exponential distribution, the mean lead time is equal to \( E(X) \). This result is a consequence of the Markovian property of the exponential distribution, such that the conditional distribution of the time remaining in the preclinical state (lead time) is the same as the unconditional distribution of the total time spent in that state.

**2.4.2.2 Natural history of disease in a dynamic population**

In the above section, how the standard epidemiology descriptions of a disease could be derived from the joint distribution of age and disease-state sojourn time with assumption of no deaths, immigration, or emigration in the absence of screening is shown. In order to obtain quantitative predictions about the effect of a screening program, it is necessary to develop a mathematical model which relates the dynamic statistical
behavior of a population’s disease natural history characteristics to the characteristics
of the screening program. In this section, such a model [12] is described.

At any instant time $t$, each individual has an associated attribute vector $Z=(Y,X)$
and age $A(t)$. The attribute vector doesn’t change with time, but age does, i.e.
$A(t+s)=A(t)+s$. Let $f_{Z,A}(t)$ donate the density of $(Z, A(t))$ over the study population,
and $N(t)$ the total population size. Then the expected number of individuals at time
$t$ whose $(Z, A(t))$ value falls within the cell $(z, z + \Delta z) \times (a, a + \Delta a)$ is expressed as:

$$n^{(t)}(z, a) \Delta z \Delta a \equiv f_{Z,A}^{(t)}(z, a) N(t) \Delta z \Delta a.$$ 

If an individual is chosen from the study population and subjected to a screening
examination, then there exist certain probabilities, i.e. false negative rate and false
positive rate. Most realistic models for tumor detectability can be synthesized by the
following model.

$$\phi = \begin{cases} 
0 & \text{if } a < y \\
1 - \theta & \text{otherwise}. 
\end{cases}$$

Where $\phi(z, a)$ is the detection probability that a women will screen positively if her
disease vector is $z=(Y,X)$ and her age at the time of screening is $a$. that is, sensitivity. 
In this case, there is a zero false positive rate and a constant false negative rate $\theta$. 
Thus a number of the stratum of individuals with $Z=z$ and $A(t)=a$ (at time $t$) will
be screened in the interval $(t, t+\delta)$ with probability $r(z,a,t)\delta$, where $r(z,a,t)$ is the
screening rate at time $t$ in stratum $Z=z, A(t)=a$. Having been chosen for screening,
there is a probability $\phi(z,a)$ that this person will screen positively. True positives are
removed from the population and never return. Thus, the equation below describes
the temporal evolution of $n^{(t)}(z, a)$.

$$\frac{\partial n^{(t)}}{\partial t}(z, a + t) = M(t) f_{Z,A}^{(t)}(z, a + t) - \Psi(z, a + t) n^{(t)}(z, a + t), \quad 0 < a + t < Y + X.$$
Here $M(t)$ is the immigration rate. $g^{(t)}_{Z,A}(\cdot)$ is the joint density of $Z$ and $A$ among those who immigrate at time $t$. and $\Psi^{(t)}(z,a) = r(z,a,t)\phi(z,a) + d(z,a,t)$, where $r(z,a,t)$ is the same as above. $\phi(z,a)$ is the probability of a positive screen. and $d(z,a,t)$ is the death rate at time $t$. Hence $\Psi^{(t)}(z,a)$ is the instantaneous net rate of removal of individuals from the stratum $Z=Z_A(t)$. By differentiation, the solution $n^{(t)}(z,a)$ of the above equation can be expressed in terms of $M(\cdot), g^{(t)}_{Z,A}(\cdot), \Psi^{(t)}(\cdot)$, and $N^{(0)}(z,a-t) = N(0)g^{(0)}_{Z,A}(z,a)$ [12]. This clearly shows how the natural history of a progressive chronic disease such as cancer, is depicted by the joint distribution of the attribute factors such as migration and death. Also the joint density of $Z$ and $A$ in the study population can be expressed as follows:

$$f_{Z,A}^{(t)}(z,a) = \frac{n^{(t)}(z,a)}{\int \int n^{(t)}(z,a)dzda}.$$ 

Thus, the preclinical prevalence $P(t)$ and incidence $I(t)$ at time $t$ are, respectively.

$$P(t) = \frac{\int_{0}^{\infty} \int_{0}^{a} \int_{a-x}^{\infty} n^{(t)}(y,x,a)dydxda \int_{0}^{\infty} \int_{0}^{a} \int_{a-x}^{\infty} n^{(t)}(y,x,a)dydxda}{n^{(t)}}$$

$$I(t) = \frac{\int_{0}^{\infty} \int_{0}^{a} n^{(t)}(y,a-y,a)dyda \int_{0}^{\infty} \int_{0}^{a} n^{(t)}(y,x,a)dydxda}{n^{(t)}}.$$ 

Where $n^{(t)}$ is the size of the population at time $t$.

Mean preclinical duration, mean lead time, and mean delay time are, respectively.

$$E(X|Y < \infty) = \frac{\int_{0}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} xn^{(t)}(y,x,a)dydxda \int_{0}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} n^{(t)}(y,x,a)dydxda}{\int_{0}^{\infty} \int_{0}^{\infty} \int_{0}^{\infty} n^{(t)}(y,x,a)dydxda}$$

$$E(L) = \frac{\int \int_{Y < a \leq Y + X} (y + x - a)n^{(t)}(z,a)r(z,a,t)\phi(z,a)dzda \int \int_{Y < a \leq Y + X} n^{(t)}(z,a)r(z,a,t)\phi(z,a)dzda}{\int \int_{Y < a \leq Y + X} n^{(t)}(z,a)r(z,a,t)\phi(z,a)dzda}$$
A screening program causes the joint distribution of $Z$ and $A$ to change over time, therefore, it results in the above characteristics changing over time.

### 2.4.2.3 Estimation of disease natural history

The model introduced in the above two sections is a basic description for the natural history of a progressive disease. Over a population of individuals the joint frequencies of $(Y, X, A)$ at any time $t$ can be represented by a probability distribution $f_{YXA}(y, x, a)$. This function represents a statistical description of the natural history of the disease in the population of interest. Characteristic descriptors of the disease process such as preclinical prevalence, clinical incidence, average preclinical duration, mean lead time and delay time can be computed using the joint distribution.

In this section, the method of estimating $f_{YXA}(y, x, a)$ are described [13]. This method uses the data more efficiently, can extract information unattainable by the epidemiologic method, and can include estimation of parameters such as false negative rate in the screening model.

The method of interest assumed the following: all participants in the study satisfy the condition $Y+X > A$ at $t=0$; each individual is screened at most once; only those subjects who clinically surface with $T$ time units give exact information on $X+Y$; false positive rate is zero, and false negative rate ($\theta$) is a constant throughout the preclinical state; death by other causes and other losses to follow-up are independent of the disease process. Furthermore, there is no age-cohort effect.

Whether a parametric or a nonparametric approach is used, the likelihood of the data is needed. With the possibility of loss to follow-up, the observed data for each
individual can be represented by a vector \( v = (d, b, s, a, c) \), with \( d \) the age at entry into the study; \( b \) the age at screening; \( s \) the result of screening (i.e., \( s = 1 \) as screening is positive, and \( s = 0 \) as screening is negative); \( a \), the age at clinical progression or censoring; and \( c \), the censoring indicator variable (i.e., \( c = 1 \) as \( a \) is censored, and \( c = 0 \) as \( a \) is uncensored). The description and contribution of four classes of \( v \)-vector to the overall likelihood are described below. \( L(v) \) represents the likelihood, and \( \theta \) is the false negative rate.

In the parametric approach, when \( s = 0, c = 0 \), that is, a negative screen with exact knowledge of age at clinical progression.

\[
L(v) = P_r(\text{negative screening, } A = a | A > d) \\
= \{ Gf_{Y | a}(b | a) + \theta Ff_{Y | a}(b | a) \} f_A(a | A > d).
\]

When \( s = 0, c = 1 \), that is, a negative screen with censored age at clinical progression.

\[
L(v) = P_r(\text{negative screening, } A > a | A > d) \\
= \{ GG_{Y | a}(b | a) + \theta FG_{Y | a}(b | a) \} G_A(a | A > d).
\]

When \( s = 1, c = 0 \), that is, a positive screen with exact knowledge of age at clinical progression.

\[
L(v) = P_r(\text{positive screening, } A = a | A > d) \\
= (1 - \theta) Ff_{Y | a}(b | a) f_A(a | A > d).
\]

When \( s = 1, c = 1 \), that is, a positive screen with censored age at clinical progression.

\[
L(v) = P_r(\text{positive screening, } A > a | A > d) \\
= (1 - \theta) FG_{Y | a}(b | a) G_A(a | A > d).
\]

Where

\[
Ff_{Y | a}(b | a) = P_r(Y \leq b | A = a) = 1 - Gf_{Y | a}(b | a)
\]
\[ FG_{Y|A}(b|a) = P_r(Y \leq b | A > a) = 1 - FG_{Y|A}(b|a). \]

Therefore, the \( i \)th individual generates data \( v_i \), and the overall likelihood for \( n \) individuals is

\[ L(v_1, v_2, \ldots, v_n) = \prod_{i=1}^{n} L(v_i) \]

The likelihood is to be maximized with respect to the unknown parameters.

In the nonparametric approach, the joint distribution is described by a finite collection of probabilities associated with regions in the plane. Let \( V_m, m = 1, 2, \ldots, M \leq 3n \), be the ordered, distinct values in the set

\[ \cup_y, \cup_f, \cup_a. \]

and define \( I_m = (V_{m-1}, V_m), m = 1, 2, \ldots, M+1 \). Let \( P(j,k) = P_r(Y \in I_j, A \in I_k) \),\( k = 1, 2, \ldots, M+1, j = 1, 2, \ldots, K \). \( n(i,j,k,s,c) \) is the number of subjects for whom \( y \in I_i, f \in I_j, a \in I + k \), and

\[ R(j,k,0,0) = \{(a,b): 1 \leq a \leq K, b = K\}, \quad R(j,k,0,1) = \{(a,b): 1 \leq a \leq b, b \geq K+1\}, \]

\[ R(j,k,1,0) = \{(a,b): 1 \leq a \leq j, b = K\}, \quad R(j,k,1,1) = \{(a,b): 1 \leq a \leq j, b \geq K+1\}. \]

The likelihood can be written as:

\[ \prod_{i=1}^{M+1} \prod_{j=i}^{M+1} \prod_{k=i}^{M+1} \prod_{s=0}^{1} \prod_{c=0}^{1} (1 - \theta)^s \sum_{(a,b) \in R(j,k,s,c)} \theta^{(1-s)w(j-a)P(a,b)} \]

\[ \frac{\prod_{i=1}^{M+1} \prod_{j=i}^{M+1} \prod_{k=i}^{M+1} \prod_{s=0}^{1} \prod_{c=0}^{1} (1 - \theta)^s \sum_{(a,b) \in R(j,k,s,c)} \theta^{(1-s)w(j-a)P(a,b)} \]}{\sum_{b=i+1}^{M+1} P(+,b)} \]

Where \( w(j-a) = 1 \) if \( j \geq a \), and \( w(j-a) = 0 \) if \( j < a \); and \( P(+,b) = \sum_{a=1}^{b} P(a,b) \). The above likelihood is to be maximized with respect to \( P(j,k)'s \) and \( \theta \).
2.4.3 The Stochastic Model of screening for progressive disease

A stochastic model was developed by Zelen [5], for early detection programs which lead to an estimate of the mean lead time as a function of observable variables. $S_p$, $S_c$ and $S_o$ are denoted respectively as a preclinical state, a clinical state, and a disease-free state. It is assumed that a person having a particular chronic disease can be regarded as being either in $S_p$ or in $S_c$, and preclinical disease eventually progresses to clinical disease if not detected and treated. Note that the probability of an individual being in $S_p$ depends on time relative to the individual, i.e., a function of when the person enters $S_p$ and the individual’s sojourn time. The probability that a person is in $S_p$ at time $t_0$ and stays there for at least $t$ additional time units is expressed as:

$$P(t_0)Q_L(t|t_0) = \int_0^{t_0} \omega(t_0 - x)Q(t + x)dx.$$

Where $P(t_0)$ is the probability of being in $S_p$ at time $t_0$, $Q_L(t|t_0)$ is the conditional probability an individual in $S_p$ at time $t_0$ stays there for at least $t$ additional time units, $\omega(t_0 - x)dx$ is the probability of a transition from $S_o$ to $S_p$ during $(t_o - x, t_o - x + dx)$, and $Q(t+x)$ is the probability of sojourn time in $S_p$.

Since $Q_L(0|t_o) = 1$.

$$Q_L(t|t_0) = \frac{\int_0^{t_0} \omega(t_0 - x)Q(t + x)dx}{P(t_0)} = \frac{\int_0^{t_o} \omega(t_0 - x)Q(t + x)dx}{\int_0^{t_0} \omega(t_0 - x)Q(x)dx}.$$  \hspace{1cm} (4.3.1)

The unconditional lead time distribution is defined as:

$$Q_L(t) = \frac{\int_0^{\infty} Q_L(t|u)P(u)du}{\int_0^{\infty} P(u)du}.$$  

Then
\[ Q_L(t) = \frac{\int_0^\infty \int_0^u \omega(u-x)Q(x+t)dxdu}{\int_0^\infty \int_0^u \omega(u-x)Q(x)dxdu} \]
\[ = \frac{\int_0^\infty \omega(x)dx \int_t^\infty Q(v)dv}{\int_0^\infty \omega(x)dx \int_0^\infty Q(v)dv} \]
\[ = \frac{\int_t^\infty Q(v)dv}{\int_0^\infty Q(v)dv} \]
\[ = \frac{\int_t^\infty Q(v)dv}{m}. \quad (4.3.2) \]

Where \( m = \int_0^\infty Q(v)dv \) is the mean sojourn time in \( S_p \).

Zelen's paper [5] indicates that if the screening program has been collecting data over a long period of time, then the appropriate lead time distribution is the unconditional distribution (4.3.2). Whereas, if the data have been collected over a short period of time, the appropriate lead time distribution is the conditional distribution (4.3.1).

Suppose the probability density function of lead time is

\[ q_L(t) = \begin{cases} 
Q(t)/m & \text{if } t \geq 0 \\
0 & \text{otherwise.}
\end{cases} \]

Then the mean lead time is

\[ L = \int_0^\infty tq_L(t)dt = \frac{m^2 + \sigma^2}{2m} = \frac{m}{2}(1 + C^2). \]

Where \( C = \sigma/m \), and \( m \) and \( \sigma^2 \) are the mean and variance of the sojourn time distribution in \( S_p \). Note that \( L > \frac{1}{2}m \) for \( \sigma^2 > 0 \). The reason is that the sojourn time probability distribution for those individuals detected in the screen is different from the population probability density function \( q(t) \). This is known as the length-biased sampling. Information on the age of incidence and the age at which a person is diagnosed in the preclinical state may be useful to assess the time gained (lead time) in diagnosing a disease using a diagnostic screening procedure [5].
Zelen defined a standardized mean age of incidence, \( m_I(t) \), for cases which are incident at time \( t \) and a standardized mean age, \( m_p(t) \), for individuals detected in \( S_p \) at time \( t \). Thus,

\[
\begin{align*}
    m_p(t) &= \int_0^\infty u P(u|t) du = \frac{\int_0^\infty u P(t, u) du}{P(t)} \\
    m_I(t) &= \int_0^\infty u I(u|t) du = \frac{\int_0^\infty u I(t, u) du}{I(t)}.
\end{align*}
\]

Where \( P(t,u)du \) and \( I(t,u)du \) are the probabilities of a person having age within the interval \( (u, u+du) \) and being in \( S_p \) at time \( t \), and entering \( S_c \) at time \( t \), respectively. As \( L \) tends to infinity, the author further derived the following:

\[
\Delta m = \lim [m_I(t) - m_p(t)] = \frac{1}{2} m (1 - C^2).
\]

Therefore,

\[
L = \frac{1}{2} m (1 + C^2) = (P/I) - \Delta m. \quad (4.3.3)
\]

where \( P \), \( I \) and \( m \) are referred to as the prevalence, incidence and mean duration of illness (sojourn time), respectively. The relationship is exact if \( m = P/I \) and is approximate if \( m \sim P/I \). The relation, given by (4.3.3), is the principal result for estimating the mean lead time associated with a one-time screening program, which is written in terms of observed variables. The author has also shown that \( m_p(t) \) is equal to \( m_I(t) \) if and only if the preclinical sojourn time is exponential.

### 2.4.4 Other Approaches

Schwartz [4] used a mathematical model of breast cancer to develop and analyze the benefits of serial screening for breast cancer by estimating the percentage of the
observed differences in five-year rates that are due to lead time bias and length-biased sampling and percentage due to earlier detection between a study group and a control group.

The model used to evaluate screening for breast cancer consists of many hypotheses such as the age specific probability that a woman develops a 0.5 cm tumor, the relationship between the time a woman has had the disease and the size of her tumor, the number of axillary lymph nodes involved, whether or not the disease has come to clinical attention in the absence of planned screens, the relationship between the size of the tumor, the number of lymph nodes involved at treatment, tumor growth rate, and the yearly threat of death from breast cancer, the age specific probability that a woman dies from causes other than breast cancer, and the false negative rate. Schwartz suggested, that in order to adjust for lead time bias, the five-year survival rate in the control group should be compared to the five-year (plus lead time) rate in the study group; to adjust for length-biased sampling, the five-year survival rate in the study group should be compared to the five-year survival rate not to women in the control group but to the five-year survival rate of women who have the same distribution of tumor growth rates as women in the study group, e.g., to the five-year survival rate of study group if they had not been screened but allowed to present clinically. To adjust for both biases simultaneously, the five-year survival (plus lead time) rate in the study group should be compared to the five-year survival rate of study group women if they had not been screened but presented clinically.

Chen and Prorok [4] presented a model for estimating a discrete form of the lead time distribution of cases detected in a one-time screening program. The procedure is based on a three-state progressive model of the natural history of disease and screening focuses on the age at entry into the preclinical disease state, the duration of preclinical state, and the age at the screening and observation. Estimation is accomplished by comparing observed incidence rates of the disease between a screened group and a control group in successive follow-up intervals after the screen. The procedure proposed in [4] extends previous direct estimation efforts to provide information on
the lead time distribution, not just the mean lead time. Application of the procedure is limited, however, since it is applicable to a one-time screening program.

A probabilistic model for disease screening applied to breast cancer was developed by Dubin [15]. He assumed, in the absence of screening, the age at diagnosis with breast cancer has a Weibull probability distribution. The parameters for this distribution were determined by fitting the observed incidence pattern in the control and refused screening groups. He showed the observed morbidity and mortality in the study (HIP) to be well fitted by the predicted values obtained from the model. Sensitivity of screening is determined to be higher in the older age groups.

Pelikan and Moskowitz [16] recently presented a model of breast cancer screening program and reported the results of this specific computer-based simulation. The model shows the effect of the screening threshold, screen sensitivity, and false negative on the cumulative mortality rate in the screened population. It suggests that factors of screen design (lead time bias and length-biased sampling) and a penalty associated with false negative assurance can result in excess mortality rates in screened population, especially in those age groups in which the incidence of rapidly growing tumors is high.

### 2.5 Summary

Estimation of parameters such as lead time, preclinical duration, and false negative rate, etc. in a screening program has been considered by several authors.

Zelen and Feinleib [5] considered the distribution of time in the preclinical state in a stochastic model, and derived an estimate of the mean lead time for the case of a single screening examination. The estimate does not depend on the form of the distribution of preclinical duration (sojourn time), but they assumed that the false negative rate of the screening test is zero.
Shapiro and Hutchison [8] and Hutchison et al. [9] made an assumption that the preclinical duration is constant, that is, all cases have the same preclinical duration, and is estimated by the ratio of prevalence at the first screen to the estimated incidence in the screened group. This approach ignores the possibility of false negative results. An alternative method proposed by them compares the observed and expected dates of diagnosis of the screened cases. Both these methods ignore the length-bias sampling issue.

Albert et al. [11,12,13] suggested an approach that examined lead time within the framework of a probability model for estimating the natural history of a disease and derived precise conditions under which the mean preclinical duration is equal to the ratio of prevalence and incidence. They discussed in more detail the prevalence to incidence ratio used by other authors, and also ignored false negatives.

Chen and Prorok [4] derived an alternative procedure for lead time estimation in a one-time screening program. The procedure is independent of assumptions about the false negative probability of the screening test and the form of the preclinical state sojourn time distribution. The procedure leads to an estimate not only for the mean lead time but, in addition, yields an estimate of a discrete form of the lead time distribution.

Shwartz [14] adapted a mathematical model to estimate the percentage of the observed difference in five-year survival rates due to lead time and length bias between women in the study and control groups. Dubin [15] also used a probabilistic model to investigate various screening parameters by fitting the number of cases diagnosed at various times and the number of deaths in various study groups over time. The preclinical duration is assumed to follow a Weibull distribution.

Although the HIP data have been analyzed by many authors, Walter and Day's approach [7] is the first which simultaneously gives estimates of the preclinical state duration, the sensitivity of the screening method, and the underlying incidence rate.
in the screened group, while also taking into account the problems of selection bias, length-biased sampling, and false negative results. Details of this approach are described in Chapter 3.

Summary of previously published results on lead time are given in table 2.1.
Table 2.1. Summary of lead time estimates of previous studies.

<table>
<thead>
<tr>
<th>Authors</th>
<th>Study</th>
<th>Selection bias</th>
<th>Length (distribution) bias</th>
<th>False negative rate</th>
<th>mean lead time estimate (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zelen and Feinleib [5]</td>
<td>HIP</td>
<td>Yes</td>
<td>Yes (any)</td>
<td>No</td>
<td>2.14</td>
</tr>
<tr>
<td>Hutchison and Shapiro [8, 9]</td>
<td>HIP</td>
<td>Yes</td>
<td>No (constant)</td>
<td>No</td>
<td>0.85</td>
</tr>
<tr>
<td>Albert et al. [11, 12, 13]</td>
<td>HIP</td>
<td>Yes</td>
<td>Yes (not specified)</td>
<td>No</td>
<td>-</td>
</tr>
<tr>
<td>Chen and Prorok [4]</td>
<td>HIP</td>
<td>Yes</td>
<td>Yes (not specified)</td>
<td>Yes</td>
<td>1.28</td>
</tr>
<tr>
<td>Shwartz [14]</td>
<td>HIP + BCDDP</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>1.9</td>
</tr>
<tr>
<td>Dubin [15]</td>
<td>HIP</td>
<td>Yes</td>
<td>Yes (Weibull)</td>
<td>Yes</td>
<td>0.4</td>
</tr>
<tr>
<td>Walter and Day [7]</td>
<td>HIP</td>
<td>Yes</td>
<td>Yes (Exponential, Step function, and Lognormal)</td>
<td>Yes</td>
<td>1.7</td>
</tr>
<tr>
<td>Pelikan and Moskowitz [16]</td>
<td>-</td>
<td>Yes</td>
<td>Yes (not specified)</td>
<td>Yes</td>
<td>-</td>
</tr>
</tbody>
</table>
Chapter 3

Materials and Methods

3.1 Data source - the National Breast Screening Study (NBSS)

The National Breast Screening Study [3] was conducted in Canada, with 15 participating screening centers across the country. The study was initiated to evaluate breast cancer screening in Canada before any decision about performing community wide screening would be made.

The NBSS is an individually randomized trial, designed to evaluate the efficacy of the combination of annual mammography, physical examination of the breast and the teaching of breast self-examination in the reduction of mortality from breast cancer in women aged 40 to 49 on entry to the study, and the combination of annual mammography over and above annual physical examination of the breast and the teaching of breast self-examination in the reduction of mortality from breast cancer in women aged 50 to 59 on entry to the study.

The NBSS involved approximately 90,000 women from 15 study centers across the country, between 40 to 59 years at entry, and satisfied the eligibility criteria: no mammography in the previous 12 months; no history of breast cancer; and no current
pregnancy. After signing an informed consent form, they were assigned to two age groups 40 to 49 years and 50 to 59 years, and were randomized in almost equal numbers to either a study group or a control group. Approximately 50,430 women age 40 to 49 years, who consented to enter the study, were randomly allocated into two groups: a screened group which was offered an annual mammogram and physical examination for up to a total of four or five screens, depending on the entry date of the woman's first screening visit; and a control group of women who had an initial physical examination only, not returning for re-screening, but were also followed-up annually by a mailed questionnaire. Approximately 39,405 women age 50 to 59 years, who consented to enter the study, were randomly allocated into either a control group or a screened group. The control group was only offered an annual physical examination alone. The screened group had an annual mammography plus an annual physical examination. Women from both groups were offered annual screening for four or five times. All women were taught breast self-examination.

Mammography and physical examination were performed independently. If an abnormality was detected, the woman was referred to the NBSS review clinic. The study surgeon played an important role in discussing the mammography findings with the study radiologist, examining the woman and deciding whether the woman required further diagnostic procedures such as routine mammogram, special views, an aspiration, or a biopsy. Once a diagnostic procedure was recommended and forwarded to the family physician, it was the responsibility of family physician to manage the patient. Physicians were encouraged to refer women who required biopsies back to the screening center [3].

After the women completed their screening schedule, direct follow-up stopped for those with no diagnosis of breast cancer. All women known to have breast cancer were followed up annually by the NBSS central office with their surgeon or other contact physician.
3.2 Study objectives

The objectives of this thesis are:

1. To obtain simultaneously point estimates of the mean duration of the detectable preclinical phase, the false negative rate of the screening test (\( \theta \)), and the underlying incidence rate (\( I \)) in the NBSS population using the Walter and Day methodology.

2. To estimate the lead time for different allocations and age groups.

3. To construct the corresponding confidence intervals for parameters lead time, \( \theta \), and \( I \) of the screening test by using the bootstrap percentile method.

4. To assess the reliability of the estimates of the parameters \( \lambda \), \( \theta \), and \( I \) and to measure the associated biases.

3.3 Walter and Day's methods used to estimate the incidence rate, the false negative rate, and the lead time

The incidence rate of disease in the population as a whole is assumed, for convenience, as a constant, denoted by \( I \), over the time prior to the use of screening and in the absence of the screening. At the time of the first screen\( (t_1) \), a certain number of cases are detected as prevalent. Immediately after \( t_1 \), the incidence rate will drop to a value below \( I \), and actually for a very short time period, the incident cases will almost entirely be comprised of persons who had false negative screening results. If the screening false negative rate is denoted by \( \theta \), then the incidence rate just after \( t_1 \) will be \( \theta I \). As time passes following \( t_1 \), new cases will develop. The incidence rate thus rises progressively until the second screen, \( t_2 \). Again, a certain number of prevalent cases are detected at the time of the second screen\( (t_2) \). Immediately after
the incident cases consist almost entirely of individuals who have had two false negative screening results, or whose disease developed between \( t_1 \) and \( t_2 \) with one false negative test. Following \( t_2 \), the incidence rises again, but the rate will be less than or equal to the rate at the same time after \( t_1 \). This process will be repeated if further screens are made. The incidence rate will eventually return to 1.

The participants who have taken screening are considered according to the number of screens they have had, and the length of time between two screens, rather than assuming that the entire population is screened simultaneously. In a screening program, there are \( n \) screens which take place at time \( t_1, t_2, \ldots, t_n \), each having a false negative rate \( \theta \), assumed constant for all screens. \( X \) is the duration or the sojourn time of detectable preclinical phase (see figure 2.1), and \( f(x) \) is defined to be the probability distribution for \( x \) in the population. Thus, under these assumptions, expression for the expected incidence rate at time \( t \) (\( I^*(t) \)), and the expected prevalence at the \( n \)th screen (\( P_n(t) \)) are derived by Walter and Day [7] as follows.

\[
I^*(t) = I \sum_{i=0}^{n} \theta^{n-i} \int_{t_i}^{t} f(x) dx \tag{3.3.1}
\]

\[
P_n(t) = (1 - \theta)I \sum_{i=1}^{n} \theta^{n-i} \int_{t_{n-1}}^{\infty} \min\{x - (t_n - t_i), t_i - t_{i-1}\} f(x) dx. \tag{3.3.2}
\]

where \( t > t_n, \ t_{n+1} = t \), and \( t_0 = -\infty \).

From equations (3.3.1) and (3.3.2), the distribution of lead time (\( g_n(t) \)) for cases found by the \( n \)th screen is shown to be

\[
g_n(t) = \frac{\sum_{i=1}^{n} \theta^{n-i} \int_{t_i + t_{n-i}}^{t_{n-i-1}} f(x) dx}{\int_{0}^{\infty} \sum_{i=1}^{n} \theta^{n-i} \int_{u+t_{n-i}}^{u+t_{n-i-1}} f(x) dx du} \tag{3.3.3}
\]

In the Walter and Day method, the predicted number of cancer cases in each of the prevalence and incidence groups can be calculated by using equations (3.3.1) and
(3.3.2), based on the assumed distributions \( f(x) \). These predicted numbers are functions of the unknown parameters being estimated. The number of both prevalent and incident cases could be observed in the follow-up of the study population. Minimizing the residual sum of squares of expected cases and observed cases (i.e., goodness-of-fit \( \chi^2 \) test), the unknown screening parameters can be estimated simultaneously. To avoid the possibility of convergence to a local minimum, a number of starting values for the parameters are used in each of the goodness-of-fit tests.

This methodology provides simultaneously 3 estimates: (1) the probability distribution \( f(x) \) of the length \( (X) \) of the detectable preclinical phase; (2) the incidence rate \( I(t) \) of the disease in the population; and (3) the screening false negative rate \( \theta \).

A special corollary of equation (3.3.3) relates to the mean lead time \( (L) \) for cases found at the first screen, this being given by the following equation:

\[
E(L) = \frac{Var(X) + E(X)^2}{2E(X)}. \tag{3.3.4}
\]

where \( E(X) \) and \( Var(X) \) are the expectation and the variance of \( X \) respectively.

The details of mathematical derivation of equation (3.3.1) and (3.3.2), and calculation procedures are shown in appendices A.2 to A.4.

Once the distribution of the length of the detectable preclinical phase is determined, the implied distribution of the lead time can be calculated. Walter and Day had considered 3 different forms of the distribution of the detectable preclinical phase:

1) The exponential distribution (a special case of the Weibull distribution):

\[
f(x) = \lambda e^{-\lambda x} \tag{3.3.5}
\]

2) The lognormal distribution:

\[
f(x) = \left[\sigma x \sqrt{2\pi}\right]^{-1} \exp\left[-\frac{(\ln x - \mu)^2}{\sigma^2}\right] \tag{3.3.6}
\]
3) An empirical step function:

\[ P(x_i \leq x < x_{i+1}) = P_i. \quad (3.3.7) \]

where \((x_i, x_{i+1})\) is the \(i\)th interval of \(x\), constituting one of the "steps" in the distribution.

The three different distributions \(f(x)\) were fitted to the data from the Health Insurance Plan of Greater New York (HIP) study. They concluded that the exponential distribution gave the best fit. The best fit distribution is determined by the goodness-of-fit \(\chi^2\) statistics. With all the necessary parameters estimated, the expected number of cases of breast cancer according to the fitted model can be calculated and the comparison to the observed number can be made.

However, the Walter and Day method does not use the maximum likelihood approach and, therefore does not calculate standard errors of the screening estimates of \(I, \theta\), and \(\lambda\). To estimate standard errors of these parameters (\(I, \theta\), and \(\lambda\)) a non-parametric method as presented below could be considered.

### 3.4 Bootstrap process

The bootstrap is a data-based simulation method for statistical inference, and computer-based method for assigning measures of accuracy to statistical estimates [17]. Suppose several independent data points \(x_1, x_2, \ldots, x_n\) (i.e., a sample) are observed, for simplicity, denoted by the vector \(x = (x_1, x_2, \ldots, x_n)\), from which the statistic of interest \(\hat{\theta}(x)\) is computed. Figure (3.1) is a schematic illustration of the bootstrap process.

The bootstrap algorithm begins by generating a large number of independent bootstrap samples \(x^*1, x^*2, \ldots, x^*K\), each of size \(n\). A bootstrap sample \(x^*k = (x_1^k, x_2^k, \ldots, x_n^k), k=1, 2, \ldots, K\), is obtained by randomly sampling \(n\) times, with replacement, from the original data points \(x_1, x_2, \ldots, x_n\). Corresponding to a bootstrap sample
Figure 3.1: Schematic Illustration of Bootstrap Process

$x^k$, a bootstrap replicate $\hat{\theta}(x^k)$, $k=1, 2, \ldots, K$, can be obtained by calculating the values of the statistics $\hat{\theta}(x)$ based on the bootstrap sample $x^k$. Finally, by using the values of bootstrap replicates $\hat{\theta}(x^1), \hat{\theta}(x^2), \ldots, \hat{\theta}(x^K)$, bootstrap estimates of standard error, biases, prediction errors, and confidence intervals, etc., for the statistics of interest $\hat{\theta}(x)$ can be obtained [17]. Suppose $\hat{\theta}(x)$ is the sample mean, for instance, the bootstrap replicate $\hat{\theta}(x^k)$ is the mean of the bootstrap sample $x^k$. The bootstrap estimate of standard error $se_{boot}$ is the standard deviation of the bootstrap replications.

$$se_{boot} = \left\{ \frac{\sum_{k=1}^{K} [\hat{\theta}(x^k) - \hat{\theta}(\cdot)]^2}{K - 1} \right\}^{1/2}.$$  \hspace{1cm} (3.4.1)

Where $\theta(\cdot) = \frac{\sum_{k=1}^{K} \hat{\theta}(x^k)}{K}$ is the mean of the bootstrap replicates.

Typical values for $K$, the number of bootstrap samples, range from 50 to 200 for standard error estimation, 50 to 200 for bias estimation, and 1,000 to 2,000 for confidence
interval estimation [17].

3.5 Bootstrap confidence intervals

An interval estimate is often more useful than just a point estimate. The point estimate and the interval estimate together describe what the best location of a parameter is, and the margin of error of that. The following sections describe how the confidence intervals are constructed using the bootstrap methods.

3.5.1 Confidence intervals based on bootstrap percentile

Suppose the data $x=(x_1, x_2, \ldots, x_n)$ are randomly sampled from an unknown distribution $F$, i.e., $F \rightarrow x=(x_1, x_2, \ldots, x_n)$. Let $\hat{\theta}$ be the estimate of a parameter of interest $\theta$, and $\hat{s}e$ be the estimated standard error of $\hat{\theta}$, based on bootstrap replications. In most situations as the sample size tends larger, the distribution of $\hat{\theta}$ becomes more and more normal with the mean being $\theta$ and the variance close to $\hat{s}^2e^2$, symbolized as $\hat{\theta} \sim N(\theta, \hat{s}^2e^2)$. Therefore, the standard confidence interval with the probability equals 1-2$, is as follows:

$$[\hat{\theta}_{lo}, \hat{\theta}_{up}] = [\hat{\theta} - Z^{(1-\alpha)} \cdot \hat{s}e, \hat{\theta} + Z^{(1-\alpha)} \cdot \hat{s}e].$$

(3.5.1)

where $\hat{\theta}_{lo}$ is lower bound and $\hat{\theta}_{up}$ upper bound. $Z^{(\alpha)}$ indicates the 100$\alpha$th percentile point of the $N(0, 1)$ distribution.

In some situations, however, the normal theory assumption in constructing the confidence interval is inappropriate, for example, when the sample size is too small, data are not normally distributed, and transformation is needed. Another approach bootstrap percentile interval can be implemented. Bootstrap percentile interval can be constructed by defining the confidence interval using the percentiles of the bootstrap
histogram (i.e., a histogram of the K bootstrap replications of \( \hat{\theta}(x^k) \), \( k = 1, 2, \ldots, K \)) without having to use the normal theory assumption.

K independent bootstrap data sets \( x^1, x^2, \ldots, x^K \) are generated and the bootstrap replications \( \hat{\theta}(x^k) \), \( k = 1, 2, \ldots, K \) are computed. Let \( \hat{\theta}^{\alpha(k)} \) be the 100\( \cdot \alpha \)th percentile of the bootstrap distribution \( \hat{\theta}(x^k) \) values, that is, \( \hat{\theta}^{\alpha(k)}_K \), the K\( \cdot \alpha \)th value in the ordered list of the K replications of \( \hat{\theta}(x^k) \) (if K\( \cdot \alpha \) is not an integer, the convention given in [17] can be used). Likewise let \( \hat{\theta}^{(1-\alpha)}_K \) be the K\( \cdot (1-\alpha) \)th value in the ordered list of the K replications of \( \hat{\theta}(x^k) \). The approximate 1-2\( \alpha \) percentile interval is

\[
[\hat{\theta}_{\%lo}, \hat{\theta}_{\%up}] = [\hat{\theta}^{\alpha}_K, \hat{\theta}^{(1-\alpha)}_K].
\] (3.5.2)

The bootstrap percentile interval has some advantages over a standard interval. First, the percentile interval can be obtained directly from the data without making any assumptions and not affected by the sample size. Second, percentile interval is transformation-respecting, that is, percentile interval for any (monotone) parameter transformation is simply the percentile interval for the parameter mapped by that transformation. Third, percentile interval is range-preserving, that is, the endpoints of the interval fall within the same allowable range as the estimated parameter [17].

3.5.2 The bias-corrected and accelerated (BC\( \alpha \)) intervals

Good bootstrap confidence intervals mean that they should closely match the exact confidence intervals in which statistical theory yields an exact answer, and should give accurate coverage probabilities in all situations. That is one of the principal goals of bootstrap theory [17]. The percentile method is not a perfect one. In some cases, e.g., the estimate might be a biased estimate and there is no suitable transformation for it, some other approaches may then be used to construct a confidence interval.

In this section, an improved version of the percentile interval, in both theory and
practice, referred to as the bias-corrected and accelerated (BC$_a$) interval is described.

The percentile interval $[\hat{\theta}_{%',lo}, \hat{\theta}_{%',up}]$ is obtained directly from the percentiles of K bootstrap replications. The BC$_a$ interval is given by percentiles of the bootstrap distribution as well, but different from those described in section 3.5.2. The percentiles used in BC$_a$ depend on two numbers $\hat{a}$ and $\hat{z}_0$, referred to as the acceleration and biased-correction respectively. The approximate $1 - 2\alpha$ BC$_a$ interval is

$$(\hat{\theta}_{lo}, \hat{\theta}_{up}) = (\hat{\theta}_K^{(\alpha_1)}, \hat{\theta}_K^{(\alpha_2)}). \quad (3.5.3)$$

where

$$\alpha_1 = \Phi(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha)})})$$
$$\alpha_2 = \Phi(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha)})}).$$

Here $\Phi(\cdot)$ is the standard normal cumulative distribution function and $z^{(\alpha)}$ is the 100$\alpha$th percentile of a standard normal distribution [17]. The BC$_a$ interval is the same as the percentile interval if $\hat{a}$ and $\hat{z}_0$ are equal to zero; if not, non-zero values of $\hat{a}$ and $\hat{z}_0$ change the upper and lower limits of the BC$_a$ interval. Detailed formulae and calculation of the values of acceleration $\hat{a}$ and biased-correction $\hat{z}_0$ are in [17]. $\hat{a}$ refers to the rate of change of the standard error of $\hat{\theta}$ with respect to the true parameter value $\theta$. $\hat{z}_0$ measures the discrepancy between the median of $\hat{\theta}_K$ and $\theta$ in normal units.

In order to construct confidence intervals, at least 1,000 bootstrap replications are needed for each parameter. Although the BC$_a$ method is better than the percentile method both in theory and in practice, it needs to compute two extra parameters
based on jackknife values of a statistic and bootstrap replications before obtaining the confidence interval. Since a jackknife value for each bootstrap sample would be needed in this computation, this requires two nested levels of bootstrap sampling. Because computing a bootstrap replication is very time-consuming in this thesis, the calculation of these extra two parameter becomes the major disadvantage of applying this method in the thesis.

3.6 Bootstrap estimate of bias

The bias of \( \hat{\theta} \), an estimate of \( \theta \), is defined as the difference between the expectation of \( \hat{\theta} \) and the value of the parameter \( \theta \).

\[
\text{bias}_F = \text{bias}_F(\hat{\theta}, \theta) = E_F(\hat{\theta}) - \theta. \tag{3.6.1}
\]

A large bias of an estimate means that the estimate is not a good one, which is usually not a desirable aspect of an estimator's performance. \( \hat{\theta} \) is a variable estimator of \( \theta \), but the variability to be overwhelmingly on either the low side or the high side is not desirable. We can use the bootstrap approach to assess the bias of any estimate \( \hat{\theta} \) [17]. The bootstrap estimate of bias is defined to be the estimate \( \text{bias}_F \)

\[
\text{bias}_F = E_F(\hat{\theta}(x^*k)) - \hat{\theta}. \tag{3.6.2}
\]

where \( \hat{\theta}(x^*k) \) is a bootstrap replicate.

K independent bootstrap samples \( x^1, x^2, \ldots, x^K \) are generated, the bootstrap replications \( \hat{\theta}(x^*k), k=1, 2, \ldots, K \) are computed, and the bootstrap expectation \( E_F(\hat{\theta}(x^*k)) \) is approximated by the average \( \hat{\theta}^* = \sum_{k=1}^{K} \hat{\theta}(x^*k)/K \). The bootstrap estimate of bias based on K replications \( \text{bias}_F \) is

\[
\text{bias}_F = \hat{\theta}^* - \hat{\theta}. \tag{3.6.3}
\]
As suggested by Efron and Tibshirani, if the ratio of the estimated bias to the standard error of the estimate is small, e.g., less than 0.25, then the bias of \( \hat{\theta} \) can be ignored [17].

The mean square error (MSE) of an estimate \( \hat{\theta} \) for \( \theta \) is \( E_F[(\hat{\theta} - \theta)^2] \), a measure of accuracy that takes into account both bias and standard error.

\[
E_F[(\hat{\theta} - \theta)^2] = se_F^2(\hat{\theta}) + bias^2_F(\hat{\theta}, \theta). \tag{3.6.4}
\]

If \( bias \) is an estimate of \( bias_F(\hat{\theta}, \theta) \), then the bias-corrected estimator is

\[
\theta' = \hat{\theta} - bias. \tag{3.6.5}
\]

Using \( bias \) equal to bootstrap bias \( bias_K = \hat{\theta}^* - \hat{\theta} \), the bias-corrected estimator becomes

\[
\theta' = 2\hat{\theta} - \hat{\theta}^*. \tag{3.6.6}
\]

Since MSE takes into account both bias and standard error, using the bias-corrected estimator may cause a larger increase in the standard error, which will result in a larger mean square error. That is, even if \( \theta' \) is less biased than \( \hat{\theta} \), it may have substantially greater standard error [17].

### 3.7 Grid search method

An alternative approach to the iterative algorithm, a grid search method, is used to examine which combination of parameter values would result in the minimum of the residual sum of squares (i.e., \( \chi^2 \)), and also to determine the confidence intervals for each of the parameters.

First, an appropriate range for each parameter is determined. Within each range, a set of points is selected. Then the chi-square values for all combinations of the values
for each parameter are calculated. The minimized chi-square is observed among these combinations, denoted by $\chi^2_{min}$. This grid search method is an alternative approach to the iterative algorithm in finding the minimum of the chi-square values.

In order to construct the confidence interval for the parameter of interest (e.g., $\lambda$), an appropriate range for $\lambda$ is chosen. Corresponding to any fixed value of $\lambda$ in this range, along with the other two parameters (e.g., $I$ and $\theta$), the minimum (i.e., a function of $\lambda$) of the residual sum of squares can be found by conducting a grid for parameter $I$ and $\theta$, denoted by $\chi^2(\lambda)$. The minimum of the residual sum of squares ($\chi^2(\lambda)$) that differs from the minimum $\chi^2_{min}$ by at most 3.84 (for 95% confidence) can be identified. The figure 3.84 is the corresponding percentile of the $\chi^2$ distribution with 1 degree of freedom. The 95% confidence interval for the parameter of interest ($\lambda$), therefore, consists of all values which result in the the minimum of the residual sum of squares ($\chi^2(\lambda)$) whose increase in chi-square relative to the minimum ($\chi^2_{min}$) is 3.84 or less.

### 3.8 Application

The methods described in previous sections are applied to the NBSS data. As described previously, a total of 89,835 eligible women were recruited into the NBSS study. These women were stratified by 10 year age groups: 40 to 49 and 50 to 59. They were randomly allocated into either the MA (mammography + physical examination) or the PE (physical examination only) allocation group. As a result of randomization, there are four subgroups. 25,214 women in the MA 40-49 group, 25,216 women in the PE 40-49 group, 19,711 women in the MA 50-59 group, and 19,694 women in the PE 50-59 group.

In this thesis, Walter and Day’s methodology is used to obtain simultaneously the point estimates of the incidence rate, the false negative rate, and the parameter of the distribution of the length of the detectable preclinical phase, for each subgroup. The data are fitted using an iterative program written in the Splus language and using
the prevalence and incidence equations (3.3.1) and (3.3.2). A numerical integration routine computes range integrals over \( f(x) \) as required. The prevalence frequencies are modeled as a function of the number of women screened, the screen sensitivity, and the sequence number of the screen, as given by equation (3.3.2). The incidence frequencies are modeled as a function of the women-months of follow-up for the group, the screening sensitivity, the history of previous negative screen, and the length of time since they occurred, as given by equation (3.3.1).

Women in the PE 40–49 group, received an initial physical examination only, and they did not return for re-screening. They were followed annually by a mail questionnaire. The data from this group is not suitable to be analyzed for the incidence rate, false negative rate, and lead time estimation and they were, therefore, excluded from the analyses.

Previous references typically have not provided standard errors or confidence intervals for these parameter estimates. Because of the constraint on the parameter of the false negative rate \( \theta \) (\( \theta \) must be in the interval (0.1)), and because of the relatively high values for its asymptotic standard error, usual normal theory approximations may be inappropriate to establish confidence intervals. Instead, the confidence intervals based on bootstrap percentile are appropriate. Thus, after getting the point estimates, we used the bootstrap percentile approach to construct confidence intervals for all these parameters.

The overall study group and overall control group were considered as random samples from a common parent population. The 25,214 women in the group MA 40-49 are considered as a random sample from the target population - all women aged 40 to 49. Then, by randomly sampling 25,214 times, with replacement, from the original data points \( x_1, x_2, \ldots, x_{25,214} \), a bootstrap sample \( x^{*k} = (x_1^*, x_2^*, \ldots, x_{25,214}^*) \) is obtained. This sampling way is repeated to generate 1,000 independent bootstrap samples \( x^{*1}, x^{*2}, \ldots, x^{*1,000} \) each of size 25,214. Then bootstrap replicates \( \hat{I}(x^{*k}), \hat{\theta}(x^{*k}), \) and \( \hat{\lambda}(x^{*k}), k = 1, 2, \ldots, 1,000 \), can be obtained by calculating the values of the
statistics $\hat{I}$ (incidence rate), $\hat{\theta}$ (false negative rate), and $\hat{\lambda}$ (parameter of distribution of the length of the detectable preclinical phase) on each bootstrap sample. Thus, 1,000 bootstrap replications based on Walter and Day's point estimate methods are obtained. Finally, 95% percentile intervals for $I$, $\theta$, and $\lambda$ can be constructed by using the methods described in section 3.5.1.

Splus language and SAS software were used for performing all statistical analyses for this thesis.
Chapter 4

Results

4.1  Point estimates of incidence rate, false negative rate, and lead time based on the NBSS data

The methodology described in section 3.3 allows the estimation of $I$ (incidence rate), $\theta$ (false negative rate), and the parameter $\lambda$ of $f(x)$ (the distribution of the length of pre-clinical detectable phase), simultaneously. The method was fitted to the NBSS data by using an iterative program written in the Splus language. Using the prevalence and incidence functions (3.3.2) and (3.3.1), a numerical integration routine computes range integrals over $f(x)$ as required. For women who attended the first screen only, prevalence at the screen and incidence rates for each of the next six years of follow-up were observed. Similarly, for women who attended the first two screens the prevalence at each screen, the incidence between the two screens and the incidence in the five years after the second screen were observed. Approximately 10 per cent of the accepters attended screening irregularly, for example, attending only the initial and fourth screen; 3 cases of breast cancer occurred in this group during the first five years of the study. Unfortunately, the detailed times of occurrence of these cases, or the women-months of follow-up for irregular attendees were not able to be determined.
For simplicity, therefore, the data from such persons have been omitted: it was felt that this is unlikely to alter substantially the conclusions presented below. The five prevalence estimates and 20 incidence estimates as shown in Tables 4.1 to 4.8, were, therefore, used. Once the data were arranged in the format as shown in Tables 4.1 to 4.8, the parameters of interest could be estimated. The last two columns in Tables 4.1 to 4.8 also show the predicted number of cancer cases in each of the two age groups (40-49 years and 50-59 years) and two allocation arms (MA - mammography and - PE control). The estimation of expected number of cancers was done assuming an exponential or Weibull distribution for the duration of the detectable preclinical phase. As shown in these tables, the expected number of prevalence or incidence calculated based on the assumed distribution did not differ from the observed cases.

Tables 4.9 and 4.10 give summary statistics arising from the fit of each distribution f(x) to the study group data (age groups and allocation arms).

Table 4.9 shows the results using the exponential distribution \( f(x) = \lambda e^{-\lambda x} \). The distribution fits satisfactorily for the age groups and allocation arms studied. The goodness-of-fit test shows a non-significant lack of fit (i.e., all p-values are greater than 0.25). The exponential function fits slightly better for the MA 50-59 group, having a smaller residual sum of squares (goodness-of-fit \( \chi^2 \) statistics) with the same number of degrees of freedom. The mean preclinical duration \( 1/\lambda \) of different groups studied varied to some extent: the smallest one is 1.51 years for the group PE 50-59, compared to 2.28 years for the MA 40-49, and 3.64 years for the MA 50-59. Note that with the exponential distribution, the mean lead time is equal to \( E(X) \), i.e., \( 1/\lambda \), as a consequence of the Markovian property of the exponential distribution. Therefore, the smallest lead time \( (1/\lambda) \) is 1.51 years for the group PE 50-59 and largest one is 3.64 years for the MA 50-59 group. The \( \hat{I}s \) (incidence) for different groups are almost identical, with 2.34 per 1,000 for the MA 40-49, 2.91 per 1,000 for the MA 50-59, and 2.72 per 1,000 for the PE 50-59. The \( \hat{d}s \) (false negative rate) estimated varied among different groups studied. The smallest estimate of \( \hat{d} \) given by the exponential function is 19.66% for the MA 40-49, compared to 27.73% for the MA 50-59, and 32.98% for
the PE 50-59.

Table 4.10 shows the results using the Weibull distribution \( f(x) = x^{\gamma-1}e^{-\lambda x^\gamma} \). Similar to the exponential distribution, the Weibull distribution also fits satisfactory for the age groups and allocation arms studied. The goodness-of-fit test shows a non-significant lack of fit (i.e., all p-values are greater than 0.25). The Weibull function fits slightly better for the MA 40-49 group, with a smaller residual sum of squares (goodness-of-fit \( \chi^2 \) statistics) with the same number of degrees of freedom. The mean preclinical duration of different groups studied varied to some extent: the smallest one is 1.69 years for the group PE 50-59, compared to 2.99 years for the MA 40-49, and 4.85 years for the MA 50-59. Note that with the Weibull distribution, the mean lead time is equal to \( (Var(X) + E(X)^2)/2E(X) \). Therefore, the smallest lead time is 2.25 years for the group MA 40-49 and largest one is 5.99 years for the group MA 50-59. The \( \hat{I} \)s (incidence) for different groups are almost identical, with 2.14 per 1.000 for the MA 40-49, 2.83 per 1.000 for the MA 50-59, and 2.91 per 1.000 for the PE 50-59. The \( \hat{\theta} \)s (false negative rate) estimated varied among different groups studied. The smallest estimate of \( \hat{\theta} \) given by the Weibull function is 4.07% for the PE 50-59, compared to 30.23% for the MA 50-59, and 48.20% for the MA 40-49.

Overall, the Weibull function fits better than the exponential function, having a smaller residual sum of squares with degrees of freedom one less than that of the exponential function. The value of the other estimators, however, are not as easily interpretable or understandable as those of the exponential function. The estimates of false negative rate for both distributions are large. The false negative rates for the HIP study and the NBSS both seem high. Since this methodology was used only in the HIP study before, this was possibly the first time that it was applied in another dataset. The properties and characteristics of the estimates of interest were not tested before. It is essential that confidence intervals for these parameters be constructed to assess the reliability. Both the exponential and the Weibull distributions fitted well in this thesis, however, literature showed the exponential had a good fit for other datasets. Compared to the Weibull distribution, the exponential distribution is much
easier to interpret and understand. Since assessing the methodology is the purpose of this thesis, a more commonly used exponential distribution was chosen in further analyses to examine the stability of these estimates. In the following section, the confidence intervals are calculated for different allocations and age groups.

4.2 Interval estimates of incidence rate, false negative rate, and lead time based on the NBSS data by the bootstrap method

The method described in section 3.5.1 was used to construct confidence intervals for the incidence rate \( l \), the false negative rate \( \theta \), the parameter of the exponential distribution \( \lambda \), which is followed by the duration of the preclinical detectable phase and the lead time \( 1/\lambda \). Figures 4.1, 4.3, and 4.5 show histograms of 1,000 bootstrap replications of \( \hat{l}(x^k) \), \( \hat{\theta}(x^k) \), and \( \hat{\lambda}(x^k) \). \( k=1, 2, \cdots, 1000 \). for the incidence rate, the false negative rate, and the parameter of exponential distribution, respectively, for the group MA 40-49. Using the bootstrap percentile interval method, the 5% and 95% percentiles of histograms of \( \hat{l} \), \( \hat{\theta} \), and \( \hat{\lambda} \) for the MA 40-49 should be roughly 0.0022 and 0.0026, 0 and 0.3701, and 0.3148 and 0.6497, respectively. Therefore, the 95% percentile intervals for \( \hat{l} \), \( \hat{\theta} \) and \( \hat{\lambda} \) are

\[
[\hat{l}_{\%10}, \hat{l}_{\%90}] = [0.0022, 0.0026] \\
[\hat{\theta}_{\%10}, \hat{\theta}_{\%90}] = [0.0000, 0.3701] \\
[\hat{\lambda}_{\%10}, \hat{\lambda}_{\%90}] = [0.3148, 0.6497].
\]

Correspondingly, the 95% percentile interval for the lead time \( 1/\hat{\lambda} \) is [1.5392, 3.1766].
The bootstrap estimates of the mean and the standard error based on 1,000 bootstrap replications \( \hat{I}(x^*) \), \( \hat{\theta}(x^*) \), and \( \hat{\lambda}(x^*) \), \( k=1, 2, \cdots, 1000 \) are 0.0028 and \( 1.4700 \times 10^{-4} \), 0.1709 and 0.1173, and 0.4590 and 0.0860, respectively. They could be used to generate the normal distribution. Figures 4.2, 4.4, and 4.6 show histograms of 1,000 random variables drawn from the normal distributions based on bootstrap standard errors of bootstrap replications of \( \hat{I}(x^*) \), \( \hat{\theta}(x^*) \), and \( \hat{\lambda}(x^*) \), \( k=1, 2, \cdots, 1000 \), that is, from \( N(0.0023, (1.4700 \times 10^{-4})^2) \), \( N(0.1966, (0.1173)^2) \), and \( N(0.4381, (0.0860)^2) \), respectively. The 95% standard intervals for \( \hat{I}, \hat{\theta}, \) and \( \hat{\lambda} \) are

\[
\begin{align*}
[\hat{I}_{lo}, \hat{I}_{up}] &= 0.0023 \pm 1.96 \times 0.0002 = [0.0020, 0.0026] \\
[\hat{\theta}_{lo}, \hat{\theta}_{up}] &= 0.1966 \pm 1.96 \times 0.1173 = [-0.0333, 0.4265] \\
[\hat{\lambda}_{lo}, \hat{\lambda}_{up}] &= 0.4381 \pm 1.96 \times 0.0860 = [-0.2695, 0.6066].
\end{align*}
\]

Correspondingly, the 95% standard interval for the lead time \( 1/\hat{\lambda} \) is (-\( \infty \), \( \infty \)).

Compared to the normal histograms for different parameters in figures 4.1 to 4.6, the bootstrap histograms of \( \hat{I}(x^*) \) and \( \hat{\lambda}(x^*) \), \( k=1, 2, \cdots, 1000 \), are more symmetric and approximately normal in shape. The distribution for the false negative rate \( \hat{\theta} \) however, is rather asymmetric, with a long tail to the right. There are large discrepancies between the normal and percentile intervals for \( \hat{\theta}, \hat{\lambda}, \) and \( 1/\hat{\lambda} \), but identical for \( \hat{I} \). The lower limits of the standard intervals for \( \hat{\theta} \) and \( \hat{\lambda} \) are negative, and they do not make sense in this thesis. It shows that standard interval does not have range-preserving property. The asymmetry also represents an important part of the improvement in the coverage for percentile interval method. These are the reasons why percentile intervals were chosen in this thesis.

Figures 4.7, 4.9, and 4.11 for the group MA 50-59 show histograms of 1,000 bootstrap replications of \( \hat{I}(x^*) \), \( \hat{\theta}(x^*) \), and \( \hat{\lambda}(x^*) \), \( k=1, 2, \cdots, 1000 \), for the incidence rate, the false negative rate, and the parameter of exponential distribution, respectively. By bootstrap percentile interval method, the 95% percentile intervals for \( \hat{I}, \hat{\theta}, \) and \( \hat{\lambda} \) are
\[
\begin{align*}
[\hat{I}_{95\%}, \hat{I}_{up}] &= [0.0026, 0.0035] \\
[\hat{\theta}_{95\%}, \hat{\theta}_{up}] &= [0.0766, 0.4059] \\
[\hat{\lambda}_{95\%}, \hat{\lambda}_{up}] &= [0.2351, 0.4230].
\end{align*}
\]

Correspondingly, the 95% percentile interval for the lead time \(1/\hat{\lambda}\) is [2.3640, 4.2535].

The bootstrap means and standard errors are 0.0029 and \(1.9600 \times 10^{-4}\), 0.2599 and 0.0840, and 0.3161 and 0.0470, respectively. They are used to generate the normal distribution. Figures 4.8, 4.10, and 4.12 show that the normal approximations based on bootstrap standard errors of \(\hat{I}(x^k), \hat{\theta}(x^k), \) and \(\hat{\lambda}(x^k), k=1, 2, \ldots, 1000.\) The standard 95% normal confidence intervals for the true values \(\hat{I}, \hat{\theta}, \) and \(\hat{\lambda}\), are

\[
\begin{align*}
[\hat{I}_{lo}, \hat{I}_{up}] &= 0.0029 \pm 1.96 \times 0.0002 = [0.0025, 0.0033] \\
[\hat{\theta}_{lo}, \hat{\theta}_{up}] &= 0.2773 \pm 1.96 \times 0.0840 = [0.1087, 0.4379] \\
[\hat{\lambda}_{lo}, \hat{\lambda}_{up}] &= 0.2733 \pm 1.96 \times 0.0470 = [0.1812, 0.3654].
\end{align*}
\]

Correspondingly, the 95% standard interval for the lead time \(1/\hat{\lambda}\) is [2.7367, 5.5188].

The bootstrap histograms look roughly normal in shape for all these three parameters. For this group, the standard intervals and percentile intervals are almost exactly same for \(\hat{I}\), but different for \(\hat{\theta}, \hat{\lambda}, \) and \(1/\hat{\lambda}\).

For the group PE 50-59, histograms of 1,000 bootstrap replications of \(\hat{I}(x^k), \hat{\theta}(x^k), \) and \(\hat{\lambda}(x^k), k=1, 2, \ldots, 1000,\) are shown in figures 4.13, 4.15 and 4.17 for the incidence rate, the false negative rate, and the parameter of exponential distribution, respectively. The 95% percentile intervals for \(\hat{I}, \hat{\theta}, \) and \(\hat{\lambda}\) are
[\hat{I}_{\%lo} \cdot \hat{I}_{\%up}] = [0.0025, 0.0032]

[\hat{\theta}_{\%lo} \cdot \hat{\theta}_{\%up}] = [0.0000, 0.5554]

[\hat{\lambda}_{\%lo} \cdot \hat{\lambda}_{\%up}] = [0.4770, 1.1701].

Correspondingly, the 95% percentile interval for the lead time \(1/\hat{\lambda}\) is [0.8546, 2.0964].

The estimated bootstrap means and standard errors are 0.0027 and \(1.7900 \times 10^{-4}\), 0.2672 and 0.1444, and 0.7709 and 0.1850, respectively. Figures 4.14, 4.16, and 4.18 show the normal approximations based on the bootstrap standard errors. So the standard 95% normal confidence intervals for the true values \(\hat{I}\), \(\hat{\theta}\), and \(\hat{\lambda}\) are

\[[\hat{I}_{lo} \cdot \hat{I}_{up}] = 0.0027 \pm 1.96 \times 0.0002 = [0.0023, 0.0031]\]

\[[\hat{\theta}_{lo} \cdot \hat{\theta}_{up}] = 0.3298 \pm 1.96 \times 0.1444 = [0.0468, 0.6128]\]

\[[\hat{\lambda}_{lo} \cdot \hat{\lambda}_{up}] = 0.6614 \pm 1.96 \times 0.1850 = [0.2980, 1.0240].\]

Correspondingly, the 95% standard interval for the lead time \(1/\hat{\lambda}\) is [0.9766, 3.3557].

Compared to normal histograms for different parameters in figures 4.13 to 4.18, the bootstrap histograms of \(\hat{I}(x^k)\) and \(\hat{\lambda}(x^k)\), \(k=1, 2, \ldots, 1000\), are quite symmetric and look roughly normal in shape. The distribution for false negative rate \(\hat{\theta}\), however, is quite asymmetric, with a long tail to the right. There are large discrepancies between the normal and percentile intervals for \(\hat{\theta}\), \(\hat{\lambda}\) and \(1/\hat{\lambda}\), but identical for \(\hat{I}\).
4.3 Bias estimates of incidence rate, false negative rate, and lead time based on the NBSS data

The methods presented in section 3.6 are used to get bootstrap estimates of biases and mean square error (MSE) for each parameter. The results are shown as follows.

For the group MA 40-49, the bootstrap estimates of biases for \( \hat{I}, \hat{\theta}, \) and \( \hat{\lambda} \), that is, the differences between the bootstrap means and the point estimates, are

\[
\begin{align*}
\text{bias}_F(\hat{I}) &= \frac{\sum_{k=1}^{K} \hat{I}(x^k)/k - \hat{I}}{\hat{\sigma}} = 0.00228 - 0.0023 = -0.00002 \\
\text{bias}_F(\hat{\theta}) &= \frac{\sum_{k=1}^{K} \hat{\theta}(x^k)/k - \hat{\theta}}{\hat{\sigma}} = 0.1707 - 0.1966 = -0.0259 \\
\text{bias}_F(\hat{\lambda}) &= \frac{\sum_{k=1}^{K} \hat{\lambda}(x^k)/k - \hat{\lambda}}{\hat{\sigma}} = 0.4590 - 0.4381 = 0.0209.
\end{align*}
\]

And

\[
\left| \frac{\text{bias}_F(\hat{I})}{\hat{\sigma}} \right| = \left| \frac{0.00002}{0.000147} \right| = 0.14 < 0.25 \\
\left| \frac{\text{bias}_F(\hat{\theta})}{\hat{\sigma}} \right| = \left| \frac{0.0259}{0.1173} \right| = 0.22 < 0.25 \\
\left| \frac{\text{bias}_F(\hat{\lambda})}{\hat{\sigma}} \right| = \left| \frac{0.0209}{0.0860} \right| = 0.24 < 0.25.
\]

where \( \hat{\sigma} \) is the bootstrap standard error of the corresponding parameter. Since these ratios are all less than 0.25 and according to the rules [17], the biases can be ignored.

The mean square error of \( \hat{I}, \hat{\theta}, \) and \( \hat{\lambda} \) are

\[
\begin{align*}
MSE(\hat{I}) &= se^2 + bias^2 = 0.04 \times 10^{-8} + 2.15 \times 10^{-8} = 2.19 \times 10^{-8} \\
MSE(\hat{\theta}) &= se^2 + bias^2 = 0.1173^2 + 0.0259^2 = 0.0144.
\end{align*}
\]
\[ MSE(\hat{\lambda}) = se^2 + bias^2 = 0.0074 + 0.0004 = 0.0078. \]

The bootstrap estimates of biases for the group MA 50-59, are

\[
\begin{align*}
\hat{bias}_S(\hat{I}) &= \sum_{k=1}^{K} \hat{I}(x^{*k})/k - \hat{I} = 0.00294 - 0.0029 = 0.00004 \\
\hat{bias}_S(\hat{\theta}) &= \sum_{k=1}^{K} \hat{\theta}(x^{*k})/k - \hat{\theta} = 0.2599 - 0.2773 = -0.0174 \\
\hat{bias}_S(\hat{\lambda}) &= \sum_{k=1}^{K} \hat{\lambda}(x^{*k})/k - \hat{\lambda} = 0.3161 - 0.2748 = 0.0387.
\end{align*}
\]

And

\[
\begin{align*}
\left| \frac{\hat{bias}_S(\hat{I})}{se} \right| &= \left| \frac{0.00004}{0.000196} \right| = 0.21 < 0.25 \\
\left| \frac{\hat{bias}_S(\hat{\theta})}{se} \right| &= \left| \frac{0.0174}{0.0840} \right| = 0.21 < 0.25 \\
\left| \frac{\hat{bias}_S(\hat{\lambda})}{se} \right| &= \left| \frac{0.0387}{0.0470} \right| = 0.82 > 0.25.
\end{align*}
\]

where \(se\) is the bootstrap standard error of the corresponding parameter. That means some biases can be ignored, but bias for \(\hat{\lambda}\) cannot be ignored.

The mean square error of \(\hat{I}\), \(\hat{\theta}\), and \(\hat{\lambda}\) are

\[
\begin{align*}
MSE(\hat{I}) &= se^2 + bias^2 = 3.88 \times 10^{-8} + 1.60 \times 10^{-8} = 5.48 \times 10^{-8} \\
MSE(\hat{\theta}) &= se^2 + bias^2 = 0.0070 + 0.0014 = 0.0084 \\
MSE(\hat{\lambda}) &= se^2 + bias^2 = 0.0470 + 0.0015 = 0.0485.
\end{align*}
\]

For the group PE 50-59, the bootstrap estimates of biases are
\[ \text{bias}_F(\hat{I}) = \frac{1}{K} \sum_{k=1}^{K} \hat{I}(x^{*k})/k - \hat{I} = 0.00272 - 0.0027 = 0.00002 \]
\[ \text{bias}_F(\hat{\theta}) = \frac{1}{K} \sum_{k=1}^{K} \hat{\theta}(x^{*k})/k - \hat{\theta} = 0.2672 - 0.3298 = -0.0626 \]
\[ \text{bias}_F(\hat{\lambda}) = \frac{1}{K} \sum_{k=1}^{K} \hat{\lambda}(x^{*k})/k - \hat{\lambda} = 0.7709 - 0.6614 = 0.1095. \]

And
\[
\left| \frac{\text{bias}_F(\hat{I})}{\hat{s}\hat{e}} \right| = \left| \frac{0.00002}{0.000179} \right| = 0.11 < 0.25
\]
\[
\left| \frac{\text{bias}_F(\hat{\theta})}{\hat{s}\hat{e}} \right| = \left| \frac{-0.0628}{0.1444} \right| = 0.44 > 0.25
\]
\[
\left| \frac{\text{bias}_F(\hat{\lambda})}{\hat{s}\hat{e}} \right| = \left| \frac{0.1095}{0.1850} \right| = 0.59 > 0.25.
\]

where \( \hat{s}\hat{e} \) is the bootstrap standard error of the corresponding parameter. That means biases for \( \hat{\theta} \) and \( \hat{\lambda} \) cannot be ignored, but bias for \( \hat{I} \) can be ignored.

The mean square error of \( \hat{I} \), \( \hat{\theta} \), and \( \hat{\lambda} \) are

\[ \text{MSE}(\hat{I}) = s\hat{e}^2 + \text{bias}^2 = 3.21 \times 10^{-8} + 0.04 \times 10^{-8} = 3.25 \times 10^{-8} \]
\[ \text{MSE}(\hat{\theta}) = s\hat{e}^2 + \text{bias}^2 = 0.0208 + 0.0039 = 0.0247 \]
\[ \text{MSE}(\hat{\lambda}) = s\hat{e}^2 + \text{bias}^2 = 0.0342 + 0.0120 = 0.0464. \]

The ratios of estimated bias to standard error for most parameters in different subgroups are quite small (i.e., less than 0.25). But in some exceptional cases, further improvement in the estimate of bias [17] may be needed.
## Table 4.1: Observed and Expected Number of Cancers (MA 40-49)

<table>
<thead>
<tr>
<th>Years since start of cases</th>
<th>Type of cases</th>
<th>No. of previous negative screens</th>
<th>No. of women screened or followed-up</th>
<th>Observed No. of cases</th>
<th>Expected No. of cases Exponential</th>
<th>Expected No. of cases Weibull</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>prevalence</td>
<td>1</td>
<td>25214</td>
<td>101</td>
<td>103.9</td>
<td>98.5</td>
</tr>
<tr>
<td>1</td>
<td>prevalence</td>
<td>2</td>
<td>22301</td>
<td>35</td>
<td>44.3</td>
<td>37.7</td>
</tr>
<tr>
<td>2</td>
<td>prevalence</td>
<td>3</td>
<td>20745</td>
<td>41</td>
<td>35.6</td>
<td>43.9</td>
</tr>
<tr>
<td>3</td>
<td>prevalence</td>
<td>4</td>
<td>19634</td>
<td>47</td>
<td>33.0</td>
<td>38.5</td>
</tr>
<tr>
<td>4</td>
<td>prevalence</td>
<td>5</td>
<td>12033</td>
<td>17</td>
<td>20.2</td>
<td>21.2</td>
</tr>
<tr>
<td>0-1</td>
<td>incidence</td>
<td>1</td>
<td>25133</td>
<td>21</td>
<td>19.8</td>
<td>27.2</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>1</td>
<td>2791</td>
<td>2</td>
<td>3.7</td>
<td>3.6</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>1</td>
<td>1587</td>
<td>1</td>
<td>2.6</td>
<td>2.4</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>1</td>
<td>1264</td>
<td>3</td>
<td>2.3</td>
<td>2.2</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>1</td>
<td>1161</td>
<td>0</td>
<td>2.3</td>
<td>2.2</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>1</td>
<td>1161</td>
<td>2</td>
<td>2.4</td>
<td>2.3</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>2</td>
<td>22266</td>
<td>15</td>
<td>13.4</td>
<td>15.7</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>2</td>
<td>2210</td>
<td>2</td>
<td>2.6</td>
<td>2.3</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>2</td>
<td>1273</td>
<td>3</td>
<td>2.0</td>
<td>1.7</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>2</td>
<td>966</td>
<td>0</td>
<td>1.7</td>
<td>1.6</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>2</td>
<td>966</td>
<td>1</td>
<td>1.9</td>
<td>1.7</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>3</td>
<td>20704</td>
<td>9</td>
<td>12.0</td>
<td>12.0</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>3</td>
<td>2806</td>
<td>2</td>
<td>3.3</td>
<td>2.7</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>3</td>
<td>1729</td>
<td>0</td>
<td>2.7</td>
<td>2.3</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>3</td>
<td>1729</td>
<td>2</td>
<td>3.1</td>
<td>2.7</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>4</td>
<td>19587</td>
<td>13</td>
<td>11.3</td>
<td>10.5</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>4</td>
<td>9016</td>
<td>13</td>
<td>10.6</td>
<td>8.5</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>4</td>
<td>9003</td>
<td>9</td>
<td>14.0</td>
<td>11.6</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>5</td>
<td>12016</td>
<td>9</td>
<td>6.9</td>
<td>6.3</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>5</td>
<td>12007</td>
<td>12</td>
<td>14.0</td>
<td>11.1</td>
</tr>
</tbody>
</table>
Table 4.2: Observed and Expected Number of Cancers (PE 40-49)

<table>
<thead>
<tr>
<th>Years since start of study</th>
<th>Type of cases</th>
<th>No. of previous negative screens</th>
<th>No. of Women screened or followed-up</th>
<th>Observed No. of cases</th>
<th>Expected No. of cases Exponential</th>
<th>Expected No. of cases Weibull</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>prevalence</td>
<td>1</td>
<td>25216</td>
<td>57</td>
<td>56.8</td>
<td>68.4</td>
</tr>
<tr>
<td>1</td>
<td>prevalence</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>prevalence</td>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>prevalence</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>prevalence</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>0-1</td>
<td>incidence</td>
<td>1</td>
<td>25159</td>
<td>33</td>
<td>37.2</td>
<td>36.8</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>1</td>
<td>25126</td>
<td>38</td>
<td>37.2</td>
<td>36.8</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>1</td>
<td>25088</td>
<td>42</td>
<td>37.1</td>
<td>36.7</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>1</td>
<td>25046</td>
<td>44</td>
<td>37.1</td>
<td>36.6</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>1</td>
<td>25002</td>
<td>40</td>
<td>37.0</td>
<td>36.6</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>1</td>
<td>24962</td>
<td>21</td>
<td>36.9</td>
<td>36.5</td>
</tr>
</tbody>
</table>
Table 4.3: Observed and Expected Number of Cancers (MA 50-59)

<table>
<thead>
<tr>
<th>Years since start of study</th>
<th>Type of cases</th>
<th>No. of previous negative screens</th>
<th>No. of Women screened or followed-up</th>
<th>Observed No. of cases</th>
<th>Expected No. of cases</th>
<th>Expected No. of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>prevalence</td>
<td>1</td>
<td>19711</td>
<td>145</td>
<td>149.8</td>
<td>151.7</td>
</tr>
<tr>
<td></td>
<td>prevalence</td>
<td>2</td>
<td>17601</td>
<td>60</td>
<td>60.2</td>
<td>58.6</td>
</tr>
<tr>
<td></td>
<td>prevalence</td>
<td>3</td>
<td>16481</td>
<td>40</td>
<td>41.9</td>
<td>42.4</td>
</tr>
<tr>
<td></td>
<td>prevalence</td>
<td>4</td>
<td>15695</td>
<td>48</td>
<td>36.9</td>
<td>36.0</td>
</tr>
<tr>
<td></td>
<td>prevalence</td>
<td>5</td>
<td>8581</td>
<td>19</td>
<td>19.9</td>
<td>19.0</td>
</tr>
<tr>
<td>0-1</td>
<td>incidence</td>
<td>1</td>
<td>19566</td>
<td>16</td>
<td>20.7</td>
<td>22.1</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>1</td>
<td>1949</td>
<td>7</td>
<td>2.9</td>
<td>2.9</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>1</td>
<td>1149</td>
<td>1</td>
<td>2.1</td>
<td>2.0</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>1</td>
<td>935</td>
<td>2</td>
<td>1.9</td>
<td>1.8</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>1</td>
<td>871</td>
<td>2</td>
<td>2.0</td>
<td>1.8</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>1</td>
<td>869</td>
<td>1</td>
<td>2.1</td>
<td>1.9</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>2</td>
<td>17541</td>
<td>11</td>
<td>11.8</td>
<td>12.9</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>2</td>
<td>1838</td>
<td>3</td>
<td>2.2</td>
<td>2.2</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>2</td>
<td>835</td>
<td>0</td>
<td>1.3</td>
<td>1.2</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>2</td>
<td>674</td>
<td>1</td>
<td>1.3</td>
<td>1.2</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>2</td>
<td>673</td>
<td>1</td>
<td>1.4</td>
<td>1.3</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>3</td>
<td>16441</td>
<td>9</td>
<td>9.7</td>
<td>10.6</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>3</td>
<td>1935</td>
<td>2</td>
<td>2.2</td>
<td>2.1</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>3</td>
<td>1193</td>
<td>1</td>
<td>1.9</td>
<td>1.7</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>3</td>
<td>1192</td>
<td>2</td>
<td>2.2</td>
<td>2.0</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>4</td>
<td>15647</td>
<td>12</td>
<td>9.0</td>
<td>9.7</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>4</td>
<td>8012</td>
<td>10</td>
<td>9.0</td>
<td>8.7</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>4</td>
<td>8002</td>
<td>5</td>
<td>12.4</td>
<td>11.3</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>5</td>
<td>8562</td>
<td>3</td>
<td>4.9</td>
<td>5.2</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>5</td>
<td>8559</td>
<td>8</td>
<td>9.6</td>
<td>9.2</td>
</tr>
</tbody>
</table>
Table 4.4: Observed and Expected Number of Cancers (PE 50-59)

<table>
<thead>
<tr>
<th>Years since start of study</th>
<th>Type of cases</th>
<th>No. of previous negative screens</th>
<th>No. of Women screened or followed-up</th>
<th>Observed No. of cases</th>
<th>Expected No. of cases</th>
<th>Expected No. of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>prevalence</td>
<td>1</td>
<td>19694</td>
<td>65</td>
<td>54.6</td>
<td>54.8</td>
</tr>
<tr>
<td>1</td>
<td>prevalence</td>
<td>2</td>
<td>17251</td>
<td>34</td>
<td>31.6</td>
<td>33.8</td>
</tr>
<tr>
<td>2</td>
<td>prevalence</td>
<td>3</td>
<td>16264</td>
<td>20</td>
<td>24.8</td>
<td>28.5</td>
</tr>
<tr>
<td>3</td>
<td>prevalence</td>
<td>4</td>
<td>15424</td>
<td>13</td>
<td>25.0</td>
<td>26.9</td>
</tr>
<tr>
<td>4</td>
<td>prevalence</td>
<td>5</td>
<td>8474</td>
<td>13</td>
<td>13.7</td>
<td>14.8</td>
</tr>
<tr>
<td>0-1</td>
<td>incidence</td>
<td>1</td>
<td>19629</td>
<td>18</td>
<td>27.4</td>
<td>22.7</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>1</td>
<td>2160</td>
<td>4</td>
<td>4.4</td>
<td>4.2</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>1</td>
<td>1280</td>
<td>1</td>
<td>3.0</td>
<td>2.9</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>1</td>
<td>1071</td>
<td>2</td>
<td>2.7</td>
<td>2.6</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>1</td>
<td>1020</td>
<td>3</td>
<td>2.7</td>
<td>2.6</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>1</td>
<td>1017</td>
<td>2</td>
<td>2.7</td>
<td>2.7</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>2</td>
<td>17417</td>
<td>18</td>
<td>20.3</td>
<td>19.5</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>2</td>
<td>2011</td>
<td>8</td>
<td>3.9</td>
<td>3.8</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>2</td>
<td>926</td>
<td>0</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>2</td>
<td>755</td>
<td>4</td>
<td>1.9</td>
<td>1.9</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>2</td>
<td>751</td>
<td>2</td>
<td>2.0</td>
<td>1.9</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>3</td>
<td>16244</td>
<td>27</td>
<td>18.3</td>
<td>18.2</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>3</td>
<td>2074</td>
<td>5</td>
<td>4.0</td>
<td>4.0</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>3</td>
<td>1287</td>
<td>3</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>3</td>
<td>1284</td>
<td>1</td>
<td>3.2</td>
<td>3.1</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>4</td>
<td>15411</td>
<td>15</td>
<td>17.3</td>
<td>17.2</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>4</td>
<td>7922</td>
<td>15</td>
<td>15.1</td>
<td>15.1</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>4</td>
<td>7907</td>
<td>5</td>
<td>18.2</td>
<td>17.9</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>5</td>
<td>8461</td>
<td>14</td>
<td>9.5</td>
<td>9.5</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>5</td>
<td>8447</td>
<td>17</td>
<td>16.1</td>
<td>16.1</td>
</tr>
</tbody>
</table>
Table 4.5: Observed and Expected Number of Cancers (MA all ages combined)

<table>
<thead>
<tr>
<th>Years since start of study</th>
<th>Type of cases</th>
<th>No. of previous negative screens</th>
<th>No. of Women screened or followed-up</th>
<th>Observed No. of cases</th>
<th>Expected No. of cases Exponential</th>
<th>Expected No. of cases Weibull</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>prevalence</td>
<td>1</td>
<td>44925</td>
<td>246</td>
<td>254.3</td>
<td>251.7</td>
</tr>
<tr>
<td>1</td>
<td>prevalence</td>
<td>2</td>
<td>39902</td>
<td>95</td>
<td>103.0</td>
<td>92.5</td>
</tr>
<tr>
<td>2</td>
<td>prevalence</td>
<td>3</td>
<td>37221</td>
<td>81</td>
<td>77.0</td>
<td>84.6</td>
</tr>
<tr>
<td>3</td>
<td>prevalence</td>
<td>4</td>
<td>35329</td>
<td>95</td>
<td>70.0</td>
<td>73.5</td>
</tr>
<tr>
<td>4</td>
<td>prevalence</td>
<td>5</td>
<td>20614</td>
<td>36</td>
<td>40.6</td>
<td>41.0</td>
</tr>
<tr>
<td>0-1</td>
<td>incidence</td>
<td>1</td>
<td>44679</td>
<td>37</td>
<td>40.0</td>
<td>45.2</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>1</td>
<td>4740</td>
<td>9</td>
<td>6.6</td>
<td>6.6</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>1</td>
<td>2413</td>
<td>2</td>
<td>4.1</td>
<td>4.1</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>1</td>
<td>2199</td>
<td>5</td>
<td>4.3</td>
<td>4.2</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>1</td>
<td>2032</td>
<td>2</td>
<td>4.3</td>
<td>4.2</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>1</td>
<td>2030</td>
<td>3</td>
<td>4.5</td>
<td>4.5</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>2</td>
<td>39807</td>
<td>26</td>
<td>24.5</td>
<td>26.0</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>2</td>
<td>4048</td>
<td>5</td>
<td>4.7</td>
<td>4.6</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>2</td>
<td>2108</td>
<td>3</td>
<td>3.3</td>
<td>3.2</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>2</td>
<td>1640</td>
<td>1</td>
<td>3.0</td>
<td>2.9</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>2</td>
<td>1639</td>
<td>2</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>3</td>
<td>37145</td>
<td>18</td>
<td>21.2</td>
<td>21.0</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>3</td>
<td>4741</td>
<td>4</td>
<td>5.4</td>
<td>5.1</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>3</td>
<td>2912</td>
<td>1</td>
<td>4.5</td>
<td>4.3</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>3</td>
<td>2912</td>
<td>4</td>
<td>5.3</td>
<td>5.1</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>4</td>
<td>35234</td>
<td>25</td>
<td>19.8</td>
<td>19.3</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>4</td>
<td>17028</td>
<td>23</td>
<td>19.2</td>
<td>18.0</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>4</td>
<td>17008</td>
<td>14</td>
<td>26.1</td>
<td>24.7</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>5</td>
<td>20578</td>
<td>12</td>
<td>11.5</td>
<td>11.2</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>5</td>
<td>20566</td>
<td>20</td>
<td>23.2</td>
<td>21.6</td>
</tr>
</tbody>
</table>
Table 4.6: Observed and Expected Number of Cancers (PE all ages combined)

<table>
<thead>
<tr>
<th>Years since start of study</th>
<th>Type of cases</th>
<th>No. of previous negative screens</th>
<th>No. of Women screened or followed-up</th>
<th>Observed</th>
<th>Expected Exponential</th>
<th>Expected Weibull</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>prevalence</td>
<td>1</td>
<td>44910</td>
<td>122</td>
<td>101.3</td>
<td>113.1</td>
</tr>
<tr>
<td>1</td>
<td>prevalence</td>
<td>2</td>
<td>17451</td>
<td>34</td>
<td>35.9</td>
<td>33.8</td>
</tr>
<tr>
<td>2</td>
<td>prevalence</td>
<td>3</td>
<td>16214</td>
<td>20</td>
<td>30.8</td>
<td>29.2</td>
</tr>
<tr>
<td>3</td>
<td>prevalence</td>
<td>4</td>
<td>15424</td>
<td>13</td>
<td>27.0</td>
<td>27.7</td>
</tr>
<tr>
<td>4</td>
<td>prevalence</td>
<td>5</td>
<td>8474</td>
<td>13</td>
<td>13.8</td>
<td>15.2</td>
</tr>
<tr>
<td>0-1</td>
<td>incidence</td>
<td>1</td>
<td>44785</td>
<td>51</td>
<td>71.8</td>
<td>47.3</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>1</td>
<td>27286</td>
<td>42</td>
<td>44.1</td>
<td>40.8</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>1</td>
<td>26368</td>
<td>43</td>
<td>42.9</td>
<td>44.9</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>1</td>
<td>26117</td>
<td>46</td>
<td>42.7</td>
<td>48.1</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>1</td>
<td>26022</td>
<td>43</td>
<td>42.8</td>
<td>50.5</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>1</td>
<td>25979</td>
<td>43</td>
<td>43.0</td>
<td>52.5</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>2</td>
<td>17417</td>
<td>18</td>
<td>25.6</td>
<td>18.2</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>2</td>
<td>20111</td>
<td>8</td>
<td>3.0</td>
<td>3.0</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>2</td>
<td>926</td>
<td>0</td>
<td>1.4</td>
<td>1.6</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>2</td>
<td>755</td>
<td>4</td>
<td>1.2</td>
<td>1.4</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>2</td>
<td>751</td>
<td>2</td>
<td>1.2</td>
<td>1.5</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>3</td>
<td>16244</td>
<td>27</td>
<td>22.0</td>
<td>17.0</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>3</td>
<td>2074</td>
<td>5</td>
<td>2.9</td>
<td>3.1</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>3</td>
<td>1287</td>
<td>3</td>
<td>1.8</td>
<td>2.2</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>3</td>
<td>1284</td>
<td>1</td>
<td>1.8</td>
<td>2.4</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>4</td>
<td>15411</td>
<td>15</td>
<td>19.4</td>
<td>16.1</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>4</td>
<td>7922</td>
<td>15</td>
<td>10.3</td>
<td>11.8</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>4</td>
<td>7907</td>
<td>5</td>
<td>10.5</td>
<td>13.4</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>5</td>
<td>8461</td>
<td>14</td>
<td>10.0</td>
<td>8.8</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>5</td>
<td>8447</td>
<td>17</td>
<td>10.3</td>
<td>12.6</td>
</tr>
</tbody>
</table>
Table 4.7: Observed and Expected Number of Cancers (All 40-49 combined)

<table>
<thead>
<tr>
<th>Years since start of study</th>
<th>Type of cases</th>
<th>No. of previous negative screens</th>
<th>No. of Women screened or followed-up</th>
<th>Observed No. of cases</th>
<th>Expected No. of cases</th>
<th>Expected Exponential</th>
<th>Expected Weibull</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>prevalence</td>
<td>1</td>
<td>50430</td>
<td>158</td>
<td>164.4</td>
<td>133.4</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>prevalence</td>
<td>2</td>
<td>22301</td>
<td>35</td>
<td>49.2</td>
<td>33.4</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>prevalence</td>
<td>3</td>
<td>20745</td>
<td>41</td>
<td>35.6</td>
<td>42.3</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>prevalence</td>
<td>4</td>
<td>19634</td>
<td>47</td>
<td>29.2</td>
<td>37.0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>prevalence</td>
<td>5</td>
<td>12033</td>
<td>17</td>
<td>16.6</td>
<td>19.7</td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>incidence</td>
<td>1</td>
<td>50272</td>
<td>54</td>
<td>60.1</td>
<td>54.9</td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>1</td>
<td>27917</td>
<td>40</td>
<td>37.4</td>
<td>35.9</td>
<td></td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>1</td>
<td>26675</td>
<td>43</td>
<td>38.8</td>
<td>39.8</td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>1</td>
<td>26310</td>
<td>47</td>
<td>40.6</td>
<td>43.6</td>
<td></td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>1</td>
<td>26163</td>
<td>40</td>
<td>42.2</td>
<td>46.4</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>1</td>
<td>26163</td>
<td>23</td>
<td>43.6</td>
<td>48.2</td>
<td></td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>2</td>
<td>22266</td>
<td>15</td>
<td>19.5</td>
<td>16.7</td>
<td></td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>2</td>
<td>2210</td>
<td>2</td>
<td>2.4</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>2</td>
<td>1273</td>
<td>3</td>
<td>1.6</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>2</td>
<td>966</td>
<td>0</td>
<td>1.3</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>2</td>
<td>966</td>
<td>1</td>
<td>1.4</td>
<td>1.7</td>
<td></td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>3</td>
<td>20704</td>
<td>9</td>
<td>15.0</td>
<td>12.9</td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>3</td>
<td>2806</td>
<td>2</td>
<td>2.7</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>3</td>
<td>1729</td>
<td>0</td>
<td>2.0</td>
<td>2.3</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>3</td>
<td>1729</td>
<td>2</td>
<td>2.3</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>4</td>
<td>19587</td>
<td>13</td>
<td>12.9</td>
<td>11.4</td>
<td></td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>4</td>
<td>9016</td>
<td>13</td>
<td>8.2</td>
<td>8.6</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>4</td>
<td>9003</td>
<td>9</td>
<td>10.0</td>
<td>11.6</td>
<td></td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>5</td>
<td>12016</td>
<td>9</td>
<td>7.5</td>
<td>6.8</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>5</td>
<td>12007</td>
<td>12</td>
<td>10.7</td>
<td>11.3</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.8: Observed and Expected Number of Cancers (All 50-59 combined)

<table>
<thead>
<tr>
<th>Years since start of study</th>
<th>Type of cases</th>
<th>No. of previous negative screens</th>
<th>No. of Women screened or followed-up</th>
<th>Observed No. of cases</th>
<th>Expected No. of cases</th>
<th>Expected No. of cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>prevalence</td>
<td>1</td>
<td>39405</td>
<td>210</td>
<td>196.5</td>
<td>203.6</td>
</tr>
<tr>
<td>1</td>
<td>prevalence</td>
<td>2</td>
<td>35052</td>
<td>94</td>
<td>92.9</td>
<td>93.2</td>
</tr>
<tr>
<td>2</td>
<td>prevalence</td>
<td>3</td>
<td>32745</td>
<td>60</td>
<td>71.3</td>
<td>69.4</td>
</tr>
<tr>
<td>3</td>
<td>prevalence</td>
<td>4</td>
<td>3119</td>
<td>61</td>
<td>64.8</td>
<td>65.3</td>
</tr>
<tr>
<td>4</td>
<td>prevalence</td>
<td>5</td>
<td>17055</td>
<td>32</td>
<td>35.2</td>
<td>35.8</td>
</tr>
<tr>
<td>0-1</td>
<td>incidence</td>
<td>1</td>
<td>39195</td>
<td>34</td>
<td>47.4</td>
<td>38.2</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>1</td>
<td>4109</td>
<td>11</td>
<td>7.2</td>
<td>6.2</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>1</td>
<td>2429</td>
<td>2</td>
<td>5.1</td>
<td>4.3</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>1</td>
<td>2006</td>
<td>4</td>
<td>4.7</td>
<td>3.9</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>1</td>
<td>1891</td>
<td>5</td>
<td>4.8</td>
<td>3.9</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>1</td>
<td>1886</td>
<td>3</td>
<td>5.0</td>
<td>4.1</td>
</tr>
<tr>
<td>1-2</td>
<td>incidence</td>
<td>2</td>
<td>32258</td>
<td>29</td>
<td>28.3</td>
<td>29.5</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>2</td>
<td>3849</td>
<td>11</td>
<td>5.9</td>
<td>5.6</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>2</td>
<td>1761</td>
<td>0</td>
<td>3.5</td>
<td>3.0</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>2</td>
<td>1429</td>
<td>5</td>
<td>3.2</td>
<td>2.7</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>2</td>
<td>1424</td>
<td>3</td>
<td>3.5</td>
<td>2.9</td>
</tr>
<tr>
<td>2-3</td>
<td>incidence</td>
<td>3</td>
<td>32685</td>
<td>36</td>
<td>26.5</td>
<td>29.8</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>3</td>
<td>4009</td>
<td>7</td>
<td>5.9</td>
<td>5.8</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>3</td>
<td>2480</td>
<td>4</td>
<td>4.8</td>
<td>4.3</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>3</td>
<td>2476</td>
<td>4</td>
<td>5.5</td>
<td>4.7</td>
</tr>
<tr>
<td>3-4</td>
<td>incidence</td>
<td>4</td>
<td>31058</td>
<td>27</td>
<td>24.7</td>
<td>28.3</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>4</td>
<td>15934</td>
<td>25</td>
<td>23.4</td>
<td>23.0</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>4</td>
<td>15909</td>
<td>10</td>
<td>30.6</td>
<td>27.3</td>
</tr>
<tr>
<td>4-5</td>
<td>incidence</td>
<td>5</td>
<td>17023</td>
<td>17</td>
<td>13.5</td>
<td>15.5</td>
</tr>
<tr>
<td>5-6</td>
<td>incidence</td>
<td>5</td>
<td>17006</td>
<td>25</td>
<td>25.0</td>
<td>24.5</td>
</tr>
</tbody>
</table>
Table 4.9: Parameter Estimates Based on the Exponential Distribution

<table>
<thead>
<tr>
<th>Estimates</th>
<th>MA 40-49</th>
<th>PE 40-49</th>
<th>MA 50-59</th>
<th>PE 50-59</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (incidence)</td>
<td>0.0023</td>
<td>Not</td>
<td>0.0029</td>
<td>0.0027</td>
</tr>
<tr>
<td>θ (false negative rate)</td>
<td>0.1966</td>
<td>calculated</td>
<td>0.2773</td>
<td>0.3298</td>
</tr>
<tr>
<td>λ (parameter in the distribution)</td>
<td>0.4381</td>
<td></td>
<td>0.2733</td>
<td>0.6614</td>
</tr>
<tr>
<td>Lead time (in years)</td>
<td>2.2826</td>
<td>3.6589</td>
<td>1.5119</td>
<td></td>
</tr>
<tr>
<td>Goodness-of-fit (χ²)</td>
<td>24.5762</td>
<td>21.4044</td>
<td>23.9407</td>
<td></td>
</tr>
<tr>
<td>Degree of freedom</td>
<td>21</td>
<td>21</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>p&gt;0.25</td>
<td>p&gt;0.25</td>
<td>p&gt;0.25</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.10: Parameter Estimates Based on the Weibull Distribution

<table>
<thead>
<tr>
<th>Estimates</th>
<th>MA 40-49</th>
<th>PE 40-49</th>
<th>MA 50-59</th>
<th>PE 50-59</th>
</tr>
</thead>
<tbody>
<tr>
<td>I (incidence)</td>
<td>0.0021</td>
<td>Not</td>
<td>0.0028</td>
<td>0.0029</td>
</tr>
<tr>
<td>θ (false negative rate)</td>
<td>0.4820</td>
<td>calculated</td>
<td>0.3023</td>
<td>0.0407</td>
</tr>
<tr>
<td>λ (parameter in the distribution)</td>
<td>0.3042</td>
<td></td>
<td>0.2283</td>
<td>0.7642</td>
</tr>
<tr>
<td>C (parameter in the distribution)</td>
<td>1.4225</td>
<td></td>
<td>0.8281</td>
<td>0.6861</td>
</tr>
<tr>
<td>Lead time (in years)</td>
<td>2.2545</td>
<td>5.9981</td>
<td>2.7363</td>
<td></td>
</tr>
<tr>
<td>Goodness-of-fit (χ²)</td>
<td>19.7887</td>
<td>20.6577</td>
<td>21.4658</td>
<td></td>
</tr>
<tr>
<td>Degree of freedom</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>p-value</td>
<td>p&gt;0.25</td>
<td>p&gt;0.25</td>
<td>p&gt;0.25</td>
<td></td>
</tr>
</tbody>
</table>
Figure 4.1: Histogram of 1,000 bootstrap replications of \( \hat{I} \), incidence rates, for the MA 40-49

![Histogram of 1,000 bootstrap replications of \( \hat{I} \), incidence rates, for the MA 40-49](image1.png)

Figure 4.2: Histogram of 1,000 random numbers from the normal distribution \( N(0.0023, 2.15E-08) \) for the MA 40-49

![Histogram of 1,000 random numbers from the normal distribution \( N(0.0023, 2.15E-08) \) for the MA 40-49](image2.png)
Figure 4.3: Histogram of 1,000 bootstrap replications of $\hat{\theta}$, false negative rate, for the MA 40-49

Figure 4.4: Histogram of 1,000 random numbers from the normal distribution $N(0.197, 0.0138)$ for the MA 40-49
Figure 4.5: Histogram of 1,000 bootstrap replications of $1/\lambda$, the lead time, for the MA 40-49

![Histogram of 1,000 bootstrap replications of $1/\lambda$, the lead time, for the MA 40-49](image)

Figure 4.6: Histogram of 1,000 random numbers from the normal distribution $N(0.4381, 0.0074)$ for the MA 40-49

![Histogram of 1,000 random numbers from the normal distribution $N(0.4381, 0.0074)$ for the MA 40-49](image)
Figure 4.7: Histogram of 1,000 bootstrap replications of \( \hat{I} \) incidence rates for the MA 50-59

![Histogram of 1,000 bootstrap replications of \( \hat{I} \) incidence rates for the MA 50-59]

Figure 4.8: Histogram of 1,000 random numbers from the normal distribution \( N(0.0029, 3.88E-08) \) for the MA 50-59

![Histogram of 1,000 random numbers from the normal distribution \( N(0.0029, 3.88E-08) \) for the MA 50-59]
Figure 4.9: Histogram of 1,000 bootstrap replications of $\hat{\theta}$, false negative rate, for the MA 50-59

Figure 4.10: Histogram of 1,000 random numbers from the normal distribution $N(0.277, 0.0070)$ for the MA 50-59
Figure 4.11: Histogram of 1,000 bootstrap replications of $1/\lambda$, the lead time, for the MA 50-59

Figure 4.12: Histogram of 1,000 random numbers from the normal distribution $N(0.2748, 0.0022)$ for the MA 50-59
Figure 4.13: Histogram of 1,000 bootstrap replications of $\hat{I}$, incidence rates, for the PE 50-59

Figure 4.14: Histogram of 1,000 random numbers from the normal distribution $N(0.0027, 3.21 \times 10^{-8})$ for the PE 50-59
Figure 4.15: Histogram of 1,000 bootstrap replications of $\hat{\theta}$, false negative rate, for the PE 50-59

Figure 4.16: Histogram of 1,000 random numbers from the normal distribution $N(0.3300, 0.0208)$ for the PE 50-59
Figure 4.17: Histogram of 1,000 bootstrap replications of $1/\tilde{\lambda}$, the lead time, for the PE 50-59

![Histogram of 1,000 bootstrap replications of $1/\tilde{\lambda}$](image)

Figure 4.18: Histogram of 1,000 random numbers from the normal distribution $N(0.6614, 0.344)$ for the PE 50-59

![Histogram of 1,000 random numbers from the normal distribution $N(0.6614, 0.344)$](image)
Chapter 5

Discussion

This thesis represents an effort to apply Walter and Day’s lead time methodology to a large database (NBSS data) and to evaluate the results. The Walter and Day methodology is designed to obtain simultaneously the estimates of the duration of the detectable preclinical phase, the sensitivity of the screening test, and the underlying incidence rate in the screened group. Once the length of the detectable preclinical phase is obtained, the lead time can be derived as well. This method has its advantages of obtaining the estimates simultaneously while considering all biases such as sample selection bias, length-biased sampling, lead time bias, and false negative rate of a screening test. This thesis is conceivably the first attempt to systematically evaluate the properties of the estimated parameters by constructing the confidence intervals for those associated parameters using the bootstrap method.

5.1 Discussion of the methods

The estimation of lead time is very important in the evaluation of screening, and in order to evaluate the distribution of deaths over time in a screened population, one has to measure the potential bias introduced by lead time. Lead time is defined as the time interval between the diagnosis of a disease by screening and the time that the disease would otherwise present clinically in the absence of screening. Firstly, lead time measures how much earlier than usual a disease is discovered. Secondly
and subsequently, it also provides a correction factor in the comparison of a screened and controlled population when examining survival experience. Thirdly, estimating lead time will also help to assess the results of the NBSS study. In the following two sections, the point estimation procedures and the confidence interval procedures in constructing the parameter estimates in a screening program based on the NBSS data are discussed.

5.1.1 Advantages and limitations of point estimates of association in this thesis

The estimation of lead time and evaluation of its role in screening programs have been investigated by epidemiologists and biostatisticians since the 1960s. Shapiro and Hutchison [8] assumed that all cases have the same preclinical duration and obtained an estimate for the lower boundary of the mean lead time. Hutchison et al. [9] developed a mean lead time estimate for the case of periodic screening. They ignored the possibility of length-biased sampling and false negative results. Albert et al. [11,12,13] suggested an approach for examining lead time in a stochastic model for estimating the natural history of a disease. Zelen and Feinleib [5] derived an estimate of the mean lead time for the case of a single screening examination, ignoring false negatives as well. Chen and Prorocok [4] proposed an alternative procedure for lead time estimation in a single-shot screening program and yielded a discrete form of the lead time distribution of cases. Dubin [15] considered the above biases, using a modeling approach by fitting the number of cases diagnosed at various times and the number of deaths in various study groups over time and taking screen sensitivity to follow a negative exponential function. He assumed however, that sensitivity is zero more than two years before clinical onset and implied the sojourn time is no longer than two years. More recently, Walter and Day [7] obtained simultaneously the estimates of the probability distribution of the length of the detectable preclinical phase, the incidence rate of disease in the population, and the screen false negative rate.
Previous approaches have attempted to estimate the length of the detectable preclinical phase and the lead time for breast cancer from the HIP study by using either screening data directly through some estimation formula, or data from a screening program and perhaps other sources, to construct an intermediate model from which lead time information was then obtained. The former approach is preferable, if it can be accomplished, since it avoids the assumptions invariably associated with any intermediate model. Many of the previous analyses involved complicated modeling or simulation approaches, often containing numerous unverifiable assumptions.

In all these previous reports, randomization provided the only reliable method of eliminating sample selection bias.

Most of the previous analyses either do not specifically consider the distribution of the detectable preclinical phase, that is, assuming it to be of constant duration and ignoring the length-biased sampling, or simply estimate only the mean value. Only Zelen and Feinleib [5] and Walter and Day [7] have discussed this issue. In addition to the exponential distribution, Walter and Day also considered the lognormal distribution and step function. Both Zelen and Feinleib, and Walter and Day agreed that the exponential probability model was the best among those compared for fitting the lead time distribution. But they all assumed that all cases found at screening would eventually have presented clinically, so that prevalent and incident cases can be considered on an equal footing. In fact, in some situations, the lesions detected may well not progress to invasive disease. Under this circumstance, the analyses should be confined to the fewer incidence cases.

The literature on the sensitivity of breast cancer screening is much less extensive than that on the lead time, but is similarly diverse in its estimates. Walter and Day's estimation of the false negative rate for mammography and physical examination combined in the HIP study is 18%, with other estimates ranging from 1% to 40%. As in other previous analyses, Walter and Day's methodology assumed the sensitivity to be independent of the time already spent in the preclinical phase. that is, the simple
assumption of a constant sensitivity. In fact, an assumption of increased sensitivity (e.g., increasing with the size of tumor) is probably more realistic, as sensitivity of a screening test is affected by the underlying distribution of time spent in the detectable preclinical phase, but it inevitably involves further parameters to be estimated in the model. For instance, a false negative rate as a function of tumor and subject age, screening modalities, and decision rules used to classify screening outcomes will be described parametrically. A sophisticated model such as a combined parametric-nonparametric approach may be used.

5.1.2 Advantages and limitations of interval and bias estimates of association in this thesis

An interval estimate is often more useful than just a point estimate. The point estimate and the interval estimate together describe the best location of a parameter, and its margin of error. The meaning of confidence interval (CI) is that over the collection of all 95% confidence intervals that could be constructed from repeated samples of size n, 95% of the time, a random interval will contain the true value of parameter. Previous reports usually did not provide standard errors or confidence regions for their parameter estimates in the screening program.

As described in section 3.5, the standard confidence intervals are almost always constructed to be equal-tail, that is, the miscoverage on each side must be $\alpha$, rather than just an overall coverage of $1-2\alpha$, because the normal distribution is assumed for the parameter estimate. This forces the interval to be symmetric. The bootstrap percentile intervals are based on the percentiles of the histogram of bootstrap replications which are directly from the data, without making any normal theory assumptions. As a consequence the resulting intervals can be asymmetric about the point estimations. This asymmetry represents an important part of the improvement in the coverage [17]. Through the use of bootstrapping, a more accurate interval can be obtained. When the sample size "n" approaches infinity, the bootstrap and standard intervals
cover each other, but in some situations the bootstrap may make substantial corrections. The inferential accuracy of the interval estimates can be improved significantly by these corrections [17].

Sometimes a transformation technique is used by statisticians to transform a highly skewed distribution to one that is more (approximately) normal. The bootstrap percentile method can be thought of as an algorithm for automatically incorporating such a transformation. The advantage of the percentile method is that the exact transformation method is not required to be specified, that is, such a transformation is assumed to exist. The standard method, however, requires a different transformation like the natural logarithm for each parameter of interest to be known. The bootstrap percentile method therefore, is referred to as transformation-respecting; the percentile interval for any (monotone) parameter transformation is simply the percentile interval for the parameter mapped by that transformation [17]. In some situations, there is a constraint for parameters on the values they can take. For instance, the value of the false negative rate must be in the interval [0,1]. Obviously it would be desirable if a confidence procedure always produced intervals within the allowable range. Such an interval is called range-preserving intervals [17]. The bootstrap percentile interval is range-preserving, since the endpoints of the interval are obtained from bootstrap replications which are subject to the same restriction as the parameter. In contrast, the standard interval need not be range-preserving. Confidence procedures that are range-preserving tend to be more accurate and reliable [17].

For purposes of this thesis, assuming the DPCP follows an exponential distribution, the 95% bootstrap percentile intervals for the lead time were calculated. They are [1.3392, 3.1766] for the MA 40-49, [2.3640, 4.2535] for the MA 50-59, and [0.8546, 2.0964] for the PE 50-59. The 95% bootstrap percentile intervals for the incidence rates are [0.0022, 0.0026], [0.0026, 0.0035], and [0.0025, 0.0032] for the group MA 40-49, MA 50-59 and PE 50-59 respectively. The 95% bootstrap percentile intervals for the false negative rates are [0.0000, 0.3701] for the group MA 40-49, [0.0766, 0.4059] for the MA 50-59, and [0.0000, 0.5554] for the PE 50-59. The length \((\hat{\theta}_{up} - \hat{\theta}_{lo})\)
and the shape \( ((\hat{\theta}_{up} - \hat{\theta})/(\hat{\theta} - \hat{\theta}_{lo})) \) of the interval \((\hat{\theta}_{lo}, \hat{\theta}_{up})\). measure the width of the interval and the asymmetry of the interval about the point estimation. The bootstrap percentile intervals and standard intervals for incidence rates of different subgroups are almost identical, not only in the length but also in the shape. Discrepancies exist between percentile and standard intervals for mean lead time and false negative rates. The standard intervals are slightly wider than the percentile intervals for most parameters in different subgroups. The bootstrap distributions are quite asymmetric for the false negative rates especially in the MA 40-49 and PE 50-59, both with long tails to the right.

The percentile method is not a perfect one. If the point estimate is biased upward or downward, the \( BC_a \) method described in section 3.5.2 and other approaches [17] automatically correct for the bias in the estimation. This is one of the advantages over the percentile method. Another advantage of the \( BC_a \) interval is its handling transformations. The percentile method is transformation-respecting and range-preserving. The standard method is neither, while the \( BC_a \) method is both as well.

Before obtaining the \( BC_a \) interval, the values of acceleration and biased-correction which are computed using jackknife values of a statistic and bootstrap replications. The \( BC_a \) interval is difficult to use because it requires a jackknife value for each bootstrap sample. This implies two nested levels of bootstrap sampling. Since it is time-consuming to compute bootstrap replications for a sample as large as the NBSS sample, the bootstrap percentile interval was chosen to construct the confidence intervals for the parameters in the screening program.

A large bias of an estimate indicates that the estimate is not a good one, which is usually not a desirable aspect of an estimator's performance. The variability to be overwhelmingly on either the low side or the high side of the true value of the parameter is undesirable [17]. In this thesis, the bootstrap method is also used to assess the bias of estimates and to see how reliable these estimates are for parameters \( \lambda, \theta, \) and \( I \). The results showed that the ratios of estimated bias to standard error
for most parameters in different subgroups are quite small (less than 0.25). In this situation, bias is of less concern. Otherwise, further improvement in the estimate is needed. For instance, increasing the number of bootstrap replications (approaching to infinity) may be needed.

The number of iterations depends on the data itself, the initial assigned values of the parameters, the incremental sizes of the parameters, and the convergence criterion. In this thesis, for the point estimates, it is observed that the estimates of parameters varied among different groups studied and could have wide ranges. If a big incremental size for the parameters is chosen, and a lower convergence criterion is selected, then some goodness-of-fit tests for the bootstrap replicates show significant lack of fit. In order to avoid these problems, smaller initial values of each parameter, a smaller incremental size, and a higher convergence criterion are chosen in this thesis. As a result, it increases the number of iterations, and therefore increases the computing time required. When computing the point estimates, some datasets needed less than 100 iterations, but some more than 200. Computing one bootstrap replication using the Splus language requires 30 minutes on average, therefore, constructing a confidence interval based on 1,000 bootstrap replications needs a month. Approximately three to four months are required to construct all the confidence intervals in this thesis, which is one of the limitations of this thesis.

This thesis suggested that the Walter and Day lead time estimation methodology can be used in the evaluation of screening. It estimates parameters of a screening program simultaneously while considering biases such as sample selection bias, length-bias sampling, lead time bias, and false negative results of a screening test. The false negative estimates obtained using Walter and Day's method are similar to what was reported by the NBSS investigators. Confidence intervals generated for the lead time and the incidence rate by the bootstrap method are relatively narrow and, therefore, precise compared to wider confidence interval around the false negative rate. This is probably because of the assumption of constant (with time) false negative rate. We have good reasons to believe that false negatives in the NBSS decreased with
time. This is part because after the first screen there were earlier mammograms to compare with, but also due to improvements in both radiologist reading skills and technical improvements in mammography incorporated in the NBSS. Therefore, questions relating to precision of the false negative rate estimated by Walter and Day's method remain. Future research is necessary to improve the estimation of the false negative rate.

5.2 Direction for future research

Since false negative rates resulting from the Walter and Day methodology based on the NBSS data and the HIP study were too high, and the preliminary results of the National Breast Screening Study suggest little or no benefit for mammography in reducing mortality from breast cancer [3], determination of lead time will help to assess whether it is either due to the methodology itself which didn't consider the possibility of increasing sensitivity with tumor growth or due to a failure (or relative insensitivity) of the screening process. Since the estimates obtained based on the exponential or the Weibull distribution for DPCP, especially for false negative rate and lead time, are quite different, which model has best fit for the NBSS data from view of the overall fit and meaningful parameter estimates needs to be further researched. A greater understanding of the detectable preclinical phase (DPCP), a greater knowledge of whether the histological lesions detected in the screening will eventually progress to invasive cancer, and a greater evaluation of increased sensitivity with time will help in assessing the estimation of the lead time.

Since there are many interval procedures that are a substantial improvement over the percentile method in both theory and practice, the algorithm for computing the point estimates in this thesis must be improved to be more efficient, in order to compare other approaches.
Bibliography


Appendix A.1

Key terms

CLINICAL BREAST CANCER (INCIDENCE) - cases of breast cancer diagnosed through usual medical care independent of the screening program

PRECLINICAL BREAST CANCER (PREVALENCE) - cases of breast cancer suspected at screening either by mammography, physical examination, or both, and subsequently confirmed by biopsy performed as a result of screening

DETECTABLE PRECLINICAL PHASE (DPCP, X) - time interval between the time a disease becomes detectable by screening and the time a disease presents clinically in the absence of screening (This term was also referred as sojourn time in the literature)

LEAD TIME (L) - the time interval between diagnosis by screening and the time disease presents clinically in the absence of screening

DETECTABLE DELAY TIME (D) - time interval between the time disease becomes detectable and diagnosis by screening

I - the underlying incidence rate of disease in the population over the time prior to the use of screening and in the absence of screening

θ - the false negative rate

I*(t) - the expected incidence rate at time t

P_n - the expected prevalence at the nth screen
Appendix A.2

Expressions for the post-screening incidence: two screens take place at times $t_1$ and $t_2$

The incidence rate of disease in the population as a whole is assumed to be constant, denoted by $I$, over the time prior to the use of the screening and in the absence of the screening. $f(x)$ will denote the probability density function of the duration of detectable preclinical phase $X$ during which the disease is preclinical but detectable by screening. For simplicity, $f(x)$ is assumed to be independent of $t$. The false negative rate, i.e. the probability that an individual in the preclinical state is screened negative, will be denoted by $\theta$; thus $1-\theta$ is the sensitivity. We assume 2 screens take place at times $t_1$ and $t_2$. Then the anticipated incidence rate at time $t$ ($>t_2$) consists of three components as follows.

(a) If the disease initiated before $t_1$ with two false negative tests at times $t_1$ and $t_2$, then the incidence rate can be expressed as:

$$I \theta^2 P_e(X > t - t_1) = I \theta^2 \int_{t_1}^{t_2} f(x)dx$$

(b) If the disease initiated between $t_1$ and $t_2$ with one false negative test at time $t_2$, then the incidence rate can be expressed as:

$$I \theta P_e(t - t_2 < X < t - t_1) = I \theta \int_{t_1}^{t_2} f(x)dx$$

(c) If the disease initiated after $t_2$, then the incidence rate can be expressed as:

$$I P_e(0 < X < t - t_2) = I \int_{t_1}^{t_2} f(x)dx$$
Therefore, the incidence rate at time \( t \) can be expressed as:

\[
I_2^*(t) = I \sum_{i=0}^{2} \theta^{2-i} \int_{t-i}^{t-i} f(x)dx
\]

Where \( t > t_2, t_3 = t, \) and \( t_0 = -\infty. \)

In general, if screens occur at times \( t_1, t_2, \ldots, t_n, \) the incidence rate \( I_n^*(t) \) after the \( n \)th screen is given by:

\[
I_n^*(t) = I \sum_{i=0}^{n} \theta^{n-i} \int_{t-i}^{t-i} f(x)dx \tag{3.3.1}
\]

Where \( t > t_n, t_{n+1} = t, \) and \( t_0 = -\infty. \)

To obtain the period incidence rate, these incidence rates are integrated over appropriate intervals.
Appendix A.3

Expressions for the prevalence at screening: two screens take place at times $t_1$ and $t_2$

We assume 2 screens take place at times $t_1$ and $t_2$. The anticipated prevalence rate at the 3rd screen at $t_3$ consists of three components as follows.

(a) If the disease enters the preclinical phase before $t_1$ with two false negative tests at times $t_1$ and $t_2$, then the prevalence rate can be expressed as:

$$(1-\theta)I \theta^2 \int_{t_0}^{t_1} \int_{t_3-u}^{\infty} f(x)dx du = (1-\theta)I \theta^2 \int_{t_0}^{t_1} \int_{t_3-u}^{\infty} I(x > t_3 - u)f(x)dx du$$

$$= (1-\theta)I \theta^2 \int_{t_0}^{t_1} \int_{t_3-u}^{\infty} I(u > t_3 - x)f(x)du dx$$

$$= (1-\theta)I \theta^2 \int_{t_0}^{t_1} \left\{t_1 - \max(t_3 - x, t_0)\right\}f(x)dx$$

$$= (1-\theta)I \theta^2 \int_{t_0}^{t_1} \min\{x - (t_3 - t_1), t_1 - t_0\}f(x)dx$$

(b) If the disease enters the preclinical phase between $t_1$ and $t_2$ with one false negative test at time $t_1$, then the prevalence rate can be expressed as:

$$(1-\theta)I \theta \int_{t_1}^{t_2} \int_{t_3-u}^{\infty} f(x)dx du = (1-\theta)I \theta \int_{t_1}^{t_2} \int_{t_3-u}^{\infty} I(x > t_3 - u)f(x)dx du$$

$$= (1-\theta)I \theta \int_{t_1}^{t_2} \int_{t_3-u}^{\infty} I(u > t_3 - x)f(x)du dx$$

$$= (1-\theta)I \theta \int_{t_1}^{t_2} \left\{t_2 - \max(t_3 - x, t_1)\right\}f(x)dx$$

$$= (1-\theta)I \theta \int_{t_1}^{t_2} \min\{x - (t_3 - t_2), t_2 - t_1\}f(x)dx$$
(c) If the disease enters the preclinical phase after $t_2$, then the prevalence rate can be expressed as:

\[
(1 - \theta) I \int_{t_2}^{t_3} \int_{t_3-u}^{\infty} f(x) dx du \\
= (1 - \theta) I \int_{t_2}^{t_3} \int_{0}^{\infty} I(x > t_3 - u)f(x) dx du \\
= (1 - \theta) I \int_{0}^{\infty} \int_{t_2}^{t_3} I(u > t_3 - x)f(x) du dx \\
= (1 - \theta) I \int_{0}^{\infty} \{t_3 - \max(t_3 - x, t_2)\} f(x) dx \\
= (1 - \theta) I \int_{0}^{\infty} \min\{x, t_3 - t_2\} f(x) dx
\]

Therefore, the prevalence $P_3$ at the 3rd screen at time $t_3$ is:

\[
P_3 = (1 - \theta) I \sum_{i=1}^{3} \int_{0}^{\infty} \min\{x - (t_3 - t_i), t_i - t_{i-1}\} f(x) dx
\]

Where $t_0 = -\infty$.

In general, if screens occur at times $t_1, t_2, ..., t_n$, the prevalence rate $P_n$ at the $n$th screen is given by:

\[
P_n = (1 - \theta) I \sum_{i=1}^{n} \int_{0}^{\infty} \min\{x - (t_n - t_i), t_i - t_{i-1}\} f(x) dx \tag{3.3.2}
\]

Where $t_0 = -\infty$. 
Appendix A.4

The detail procedures for results presented in Tables 4.1 to 4.8

Formulae (3.3.1) and (3.3.2) are used to generate Tables 4.1 to 4.8. Here the exponential distribution ($f(x) = \lambda e^{-\lambda x}$) is used as an example for the group MA 50-59. The estimates of $I$, $\theta$, and $\lambda$ for MA 50-59 are $I = 0.0029$, $\theta = 0.2773$, and $\lambda = 0.2733$ respectively. We assume the screens take place at times $t_1 = 0$ and $t_2 = 1$. Then the prevalence rate $P_3$ at the 3rd screen at time $t_3 = 2$ is:

\[
P_3 = (1 - \theta) I \sum_{i=1}^{3} \theta^{3-i} \int_{t_{i-1}}^{\infty} \min\{x - (t_3 - t_i), t_i - t_{i-1}\} \lambda e^{-\lambda x} dx
\]

\[
= (1 - \theta) I \left\{ \theta^2 \int_{t_{i-1}}^{\infty} \min\{x - 2, \infty\} \lambda e^{-\lambda x} dx + \theta \int_{t_{i-1}}^{\infty} \min\{x - 1, 1\} \lambda e^{-\lambda x} dx + \int_{0}^{\infty} \min\{x, 1\} \lambda e^{-\lambda x} dx \right\}
\]

\[
= (1 - \theta) I \left\{ \theta^2 \int_{t_{i-1}}^{\infty} (x - 2) \lambda e^{-\lambda x} dx + \theta \int_{t_{i-1}}^{\infty} (x - 1) \lambda e^{-\lambda x} dx + \int_{0}^{\infty} x \lambda e^{-\lambda x} dx + \int_{0}^{t_{i-1}} \lambda e^{-\lambda x} dx \right\}
\]

\[
= (1 - \theta) I \lambda^{-1} \left\{ \theta^2 e^{-2\lambda} + \theta (e^{-\lambda} - e^{-2\lambda}) + (1 - e^{-\lambda}) \right\}
\]

\[
= 0.00254
\]

Hence, the expected prevalence cases $N_p$ at the 3rd screen at time $t_3 = 2$ is as below:

\[
N_p = N \times P_3
\]

\[
= 16481 \times 0.00254
\]

\[
= 41.9
\]
If screens take place at times $t_1 = 0$ and $t_2 = 1$, then the incidence rate at time $t$ lying between the interval $(t_3, t_4)$, where $t_3 = 2$ and $t_4 = 3$, is:

$$I_2^*(t) = I \sum_{i=0}^{2} \theta^{2-i} \int_{t-i}^{t-i+1} \lambda e^{-\lambda x} dx$$

$$= I \sum_{i=0}^{2} \theta^{2-i} (e^{-\lambda (t-i+1)} - e^{-\lambda (t-i)})$$

$$= I \{ \theta^2 e^{-\lambda t} + \theta (e^{-\lambda t} - e^{-\lambda}) + (1 - e^{-\lambda (t-1)}) \}$$

Therefore, the incidence rate in the period $(2, 3)$ is:

$$I_2^*(2,3) = \int_{2}^{3} I_2(t) dt$$

$$= I \{ \theta^2 \int_{2}^{3} e^{-\lambda t} dt + \theta (\int_{2}^{3} e^{-\lambda t} dt - \int_{2}^{3} e^{-\lambda t} dt) + (\int_{2}^{3} dt - \int_{2}^{3} e^{-\lambda (t-1)} dt) \}$$

$$= I \lambda^{-1} \{ \theta^2 (e^{-2\lambda} - e^{-3\lambda}) + \theta (e^{-\lambda} - 2e^{-2\lambda} + e^{-3\lambda}) + (\lambda + e^{-2\lambda} - e^{-\lambda}) \}$$

$$= 0.0012$$

Hence, the expected incident cases $N_1$ at the time period $(2, 3)$ is:

$$N_1 = N \times I_2^*(2, 3)$$

$$= 1838 \times 0.0012$$

$$= 2.2$$