BIOSTATISTICAL APPLICATIONS IN CANCER RESEARCH

Cancer Treatment and Research

Steven T. Rosen, M.D., *Series Editor*

Goldstein, L.J., Ozols, R. F. (eds.): *Anticancer Drug Resistance. Advances in Molecular and Clinical Research.* 1994. ISBN 0-7923-2836-1.
Hong, W.K., Weber, R.S. (eds.): *Head and Neck Cancer. Basic and Clinical Aspects.* 1994. ISBN 0-7923-3015-3.
Thall, P.F. (ed): *Recent Advances in Clinical Trial Design and Analysis.* 1995. ISBN 0-7923-3235-0.
Buckner, C. D. (ed): *Technical and Biological Components of Marrow Transplantation.* 1995. ISBN 0-7923-3394-2.
Winter, J.N. (ed.): *Blood Stem Cell Transplantation.* 1997. ISBN 0-7923-4260-7.
Muggia, F.M. (ed): *Concepts, Mechanisms, and New Targets for Chemotherapy.* 1995. ISBN 0-7923-3525-2.
Klastersky, J. (ed): *Infectious Complications of Cancer.* 1995. ISBN 0-7923-3598-8.
Kurzrock, R., Talpaz, M. (eds): *Cytokines: Interleukins and Their Receptors.* 1995. ISBN 0-7923-3636-4.
Sugarbaker, P. (ed): *Peritoneal Carcinomatosis: Drugs and Diseases.* 1995. ISBN 0-7923-3726-3.
Sugarbaker, P. (ed): *Peritoneal Carcinomatosis: Principles of Management.* 1995. ISBN 0-7923-3727-1.
Dickson, R.B., Lippman, M.E. (eds.): *Mammary Tumor Cell Cycle, Differentiation and Metastasis.* 1995. ISBN 0-7923-3905-3.
Freireich, E.J, Kantarjian, H. (eds.): *Molecular Genetics and Therapy of Leukemia.* 1995. ISBN 0-7923-3912-6.
Cabanillas, F., Rodriguez, M.A. (eds.): *Advances in Lymphoma Research.* 1996. ISBN 0-7923-3929-0.
Miller, A.B. (ed.): *Advances in Cancer Screening.* 1996. ISBN 0-7923-4019-1.
Hait , W.N. (ed.): *Drug Resistance.* 1996. ISBN 0-7923-4022-1.
Pienta, K.J. (ed.): *Diagnosis and Treatment of Genitourinary Malignancies.* 1996. ISBN 0-7923-4164-3.
Arnold, A.J. (ed.): *Endocrine Neoplasms.* 1997. ISBN 0-7923-4354-9.
Pollock, R.E. (ed.): *Surgical Oncology.* 1997. ISBN 0-7923-9900-5.
Verweij, J., Pinedo, H.M., Suit, H.D. (eds.): *Soft Tissue Sarcomas: Present Achievements and Future Prospects.* 1997. ISBN 0-7923-9913-7.
Walterhouse, D.O., Cohn, S. L. (eds.): *Diagnostic and Therapeutic Advances in Pediatric Oncology.* 1997. ISBN 0-7923-9978-1.
Mittal, B.B., Purdy, J.A., Ang, K.K. (eds.): *Radiation Therapy.* 1998. ISBN 0-7923-9981-1.
Foon, K.A., Muss, H.B. (eds.): *Biological and Hormonal Therapies of Cancer.* 1998. ISBN 0-7923-9997-8.
Ozols, R.F. (ed.): *Gynecologic Oncology.* 1998. ISBN 0-7923-8070-3.
Noskin, G. A. (ed.): *Management of Infectious Complications in Cancer Patients.* 1998. ISBN 0-7923-8150-5
Bennett, C. L. (ed.): *Cancer Policy.* 1998. ISBN 0-7923-8203-X
Benson, A. B. (ed.): *Gastrointestinal Oncology.* 1998. ISBN 0-7923-8205-6
Tallman, M.S. , Gordon, L.I. (eds.): *Diagnostic and Therapeutic Advances in Hematologic Malignancies.* 1998. ISBN 0-7923-8206-4
von Gunten, C.F. (ed.): *Palliative Care and Rehabilitation of Cancer Patients.* 1999. ISBN 0-7923-8525-X
Burt, R.K., Brush, M.M. (eds): *Advances in Allogeneic Hematopoietic Stem Cell Transplantation.* 1999. ISBN 0-7923-7714-1
Angelos, P. (ed): *Ethical Issues in Cancer Patient Care* 2000. ISBN 0-7923-7726-5
Gradishar, W.J., Wood, W.C. (eds): *Advances in Breast Cancer Management.* 2000. ISBN 0-7923-7890-3
Sparano, Joseph A. (ed.): *HIV & HTLV-I Associated Malignancies.* 2001. ISBN 0-7923-7220-4.
Ettinger, David S. (ed.): *Thoracic Oncology.* 2001. ISBN 0-7923-7248-4.
Bergan, Raymond C. (ed.): *Cancer Chemoprevention.* 2001. ISBN 0-7923-7259-X.
Raza, A., Mundle, S.D. (eds): *Myelodysplastic Syndromes & Secondary Acute Myelogenous Leukemia* 2001. ISBN: 0-7923-7396.
Talamonti, Mark S. (ed.): *Liver Directed Therapy for Primary and Metastatic Liver Tumors.* 2001. ISBN 0-7923-7523-8.
Stack, M.S., Fishman, D.A. (eds): *Ovarian Cancer.* 2001. ISBN 0-7923-7530-0.
Bashey, A., Ball, E.D. (eds): *Non-Myeloablative Allogeneic Transplantation.* 2002. ISBN 0-7923-7646-3
Leong, Stanley P.L. (ed.): *Atlas of Selective Sentinel Lymphadenectomy for Melanoma, Breast Cancer and Colon Cancer.* 2002. ISBN 1-4020-7013-6
Andersson , B., Murray D., (eds.): *Clinically Relevant Resistance in Cancer Chemotherapy.* 2002. ISBN 1-4020-7200-7.
Beam, C. (ed.): *Biostatistical Issues in Cancer Research.* 2002. ISBN 1-4020-7226-0.

BIOSTATISTICAL APPLICATIONS IN CANCER RESEARCH

edited by

Craig Beam, Ph.D.
H. Lee Moffitt Cancer Center & Research Institute
University of South Florida College of Medicine
Tampa, FL, U.S.A.

SPRINGER SCIENCE+BUSINESS MEDIA, LLC

ISBN 978-1-4419-5310-0 ISBN 978-1-4757-3571-0 (eBook)
DOI 10.1007/978-1-4757-3571-0
Library of Congress Cataloging-in-Publication Data

A C.I.P. Catalogue record for this book is available
from the Library of Congress.

Printed on acid-free paper.

CONTENTS

PREFACE

Biostatistics is defined as much by its application as it is by theory. To understand modern Biostatistical thinking in any area of Medicine, it is necessary therefore to have an appreciation of the manner in which real research problems have been defined, faced and solved.

Over the past forty years Biostatistics has played an ever-increasing role in cancer research. Yet, because of this maturity, the application of Biostatistics in this specialized area of medicine is very diverse This diversity, accompanied by the often great depth of insight and development that has been achieved, makes an introduction to modern applications of Biostatistics in cancer research a formidable challenge for the novice.

The goal of this book is to provide an introduction to Biostatistical applications in modern cancer research that is both accessible and valuable to the novice cancer biostatistician or to the cancer researcher learning Biostatistics. It is hoped that accessibility and value has been achieved by placing together pairs of chapters covering a very broad swatch of problems currently encountered in cancer research. The topical areas included in this book span much of the most active areas of the application of Biostatistics to modern cancer research: survival analysis, screening, diagnostics, spatial analysis and the analysis of microarray data. In each topical area pairs of chapters are provided: one chapter giving an overview "state of the art" summary and

application taken from real research in cancer. The latter chapters are provided as something akin to "field notes" and present works in progress that should be a valuable reference and source of inspiration to the reader.

Chapters 1 and 2 cover **Statistical Models of Screening** and were contributed by Drs Marvin Zelen and Sandra J. Lee from the Harvard School of Public Health. A review and application of **Survival Analysis Methods in Cancer**, provided in Chapters 3 and 4, were contributed by Drs. John P. Klein and Mei-Jie Zhang from the Medical College of Wisconsin. The **Analysis of Microarray Data** is reviewed and an application to prostate cancer is provided in Chapters 5 and 6 by Dr. Borko D. Jovanovic and colleagues Ray C. Bergan, MD and Warren A. Kibbe, PhD from Northwestern University Medical School. Drs Linda W. Pickle and B. Sue Bell from the NCI contribute chapters 7 and 8, which provide an overview and application of statistical methods for the **Spatial Analysis of Disease.** The book then concludes with an overview of **Statistical Methods for Cancer Diagnostics** that was contributed by Dr. Alicia Toledano from Brown University. An application of these methods to mammography, written by Dr. Toledano and Benjamin Herman, SM, completes the pair.

I wish to express my deepest thanks to the contributing authors for their hard work, creativity and patience during the compilation of this book. My thanks also go to Laura Walsh, Medical Editor at Kluwer, for her guidance and even-handed oversight of this project.

Chapter 1

MODELS AND THE EARLY DETECTION OF DISEASE: METHODOLOGICAL CONSIDERATIONS

Marvin Zelen
Department of Biostatistics,
Harvard School of Public Health and
Dana-Farber Cancer Institute

Sandra J. Lee
Department of Biostatistics,
Harvard School of Public Health and
Dana-Farber Cancer Institute

1. INTRODUCTION AND BACKGROUND

There is increasing interest in using special diagnostic examinations to detect a potential disease or condition in an individual who has no signs or symptoms. The motivation behind such efforts is that earlier detection, when combined with an effective therapy, will lead to enhanced benefit. The benefit is usually reflected in a higher cure rate or longer survival. In ordinary circumstances most cancers—especially solid tumors, are usually diagnosed when an individual experiences pain or a vital organ is not functioning properly or a "lump" beneath the skin is palpated. Diagnosing a disease without these signs or symptoms often results in the disease being diagnosed in an earlier stage. Many therapies are enhanced when given to earlier disease stage patients. The primary treatment modalities for most cancer sites is surgery and/or radiation. These are essentially treatments which are targeted at localized disease. If the cancer site has metastasized, then resecting or radiating the primary tumor will not result in cures. Effective chemotherapy can be significantly enhanced if the metastatic disease consists of microscopic cells which have not seriously interfered with a vital organ.

Generally when a special diagnostic procedure is given to an individual who has no signs or symptoms of a specific disease, the exam is often referred to as a screening examination. A screening program refers to a series of scheduled screening examinations.

There are many recommended screening programs for various cancer sites. Amongst these are: breast, cervical, colorectal, ovarian and prostate cancer. There is renewed interest in screening for lung cancer. Screening programs have been used for many other chronic diseases; e.g., tuberculosis, diabetes, hypertension and coronary artery disease. National efforts are being discussed to screen high risk populations for HIV-related diseases.

In some screening programs the purpose may be to detect risk factors which may indicate that an individual is at elevated risk for a specific disease. The reason for attempting to detect these risk factors is that they may be modified; e.g., identifying individuals with hypertension, which with proper medication, blood pressure can be reduced to normal levels. The revolution in genomics is likely to result in finding many genes which identify individuals who are at a higher risk for some chronic diseases. Already genes have been found which have a role in breast and colon cancer. However without a suitable treatment, it is not at all clear, if the identification of high risk individuals for specific diseases is beneficial. The detection of prostate cancer using the prostate specific antigen (PSA) may be such an example. There is no evidence from clinical trials that therapy is enhanced with earlier detection.

Another issue is that an early detection program may identify disease which will never manifest itself during an individual's lifetime. The term "over diagnosis" is sometimes used to describe this situation. Examples of potential over diagnosis are prostate cancer and ductal carcinoma in situ (DCIS) for breast cancer. Unfortunately when an early cancer or a precursor to cancer is found it is not possible to distinguish whether the cancer will become clinical in a person's lifetime or never appear. One way to justify the potential of over diagnosis is that decisions are made using the "Minamax Principle"; i.e., minimize maximum loss. The possible losses are over-treatment or death due to disease. Clearly over treatment is a much lesser loss than death due to disease.

Mathematical models have been developed to address issues which arise in screening. Earlier investigations on screening models have been conducted by Kirch and Klein (1974), Lincoln and Weiss (1964) and Zelen and Feinleib (1969). Subsequently various statistical methods for evaluating screening schedules have been developed; cf. Albert et al. (1978), Dubin (1981), Eddy (1980, 1983), Eddy and Shwartz (1982), Kirch and Klein (1979), Prorok (1976a, b), Shwartz (1978), Shwartz and Plough (1984). In particular, Baker (1998), Baker and Chu (1990), Lee and Zelen (1998), Parmigiani (1993, 1997) and Zelen (1993) discuss optimizing examination schedules with respect to the initial age or the screening intervals.

Another aspect to the analytic approach in scheduling screening programs is estimating parameters used in the statistical models. Such parameters are the mean sojourn time of the pre-clinical state and sensitivity/positive predictive values of screening tests. These estimates have been proposed by many investigators,primarily in breast cancer screening trial settings. Chen et al. (1996, 1997), Day and Walter (1984), Duffy et al. (1995), Shen and Zelen (1999), Straatman et al. (1997) and Walter and Day (1983) used the data from breast cancer screening trials to estimate various parameters.

In addition, Etzioni et al. (1995) and Hu and Zelen (1997) addressed issues of designing early detection trials. Simulation models complement analytical models for screening programs. Knox (1973) and Habbema et al. (1984) proposed simulation models to adopt more complicated models arise in scheduling screening programs.

This chapter will mainly discuss models for early detection programs, the planning of clinical trials to evaluate the benefits of screening examinations, and problems of estimation. The accompanying chapter (Lee and Zelen) discusses the planning of public health programs for the early detection of disease. All references appear in the reference section of the Lee/Zelen paper.

2. MODELS OF THE SCREENING PROCESS:

2.1 Natural Histories

Consider that at any point in time an individual can be in one of three possible states. These are defined by:

S_o: Disease-free state. Individual has disease which cannot be detected by any examination.

S_p: Pre-clinical state. Individual has disease, but there are no signs or symptoms and the individual is unaware of having the disease. The disease may be detected by an examination.

S_c: Clinical diagnosis. Individual has been diagnosed by usual routine medical care.

The natural history of the disease is described by the paths leading and exiting these states. Examples of three natural histories are shown below.

$S_0 \rightarrow S_p \rightarrow S_c$: Progressive disease model.

$S_0 \rightarrow S_p S_p \nearrow S_c$: Progressive disease model where a subgroup may never enter the clinical state.

$S_0 \leftrightarrow S_p \rightarrow S_c$: Non-progressive disease. Preclinical state may revert back to the disease-free state.

Breast cancer, lung cancer, gastro-intestinal and genital-urinary cancers are thought to follow a progressive disease model. Prostate cancer also follows a

progressive disease model, but many men are asymptomatic who have prostate cancer and die of other causes without the disease being diagnosed. Hence a subset of cases never enter the clinical state. Cervical cancer may be a non-progressive disease in which cervical dysplasia (believed to be a pre-clinical state) may eventually disappear for some women.

Sometimes it is advantageous to add an absorbing state S_d which refers to death. Theoretically it is possible to enter the absorbing state from any other state. The usual path is $S_c \rightarrow S_d$ or $S_o \rightarrow S_d$. However if the path is $S_p \rightarrow S_d$, then it implies that a person dies without the disease ever being detected.

2.2 Length and Lead Time Bias

Survival or mortality is the usual endpoint when evaluating the benefit of a treatment for a life-threatening disease. However survival is not an appropriate endpoint for evaluating the benefit of early detection as the early detection process results in biases which make survival an inappropriate endpoint. These biases are called length-bias and lead time bias. Instead mortality or disease-specific mortality is the appropriate endpoint.

The length bias arises because individuals identified with disease in a screening program are not a random sample of people in the pre-clinical state. They tend to have longer pre-clinical sojourn times; i.e. the longer the time in the pre-clinical state, the greater the probability of being diagnosed by a screening examination. Undoubtedly the clinical course of the disease is correlated with the pre-clinical course. A short pre-clinical duration implies that the disease is aggressive whereas a long pre-clinical sojourn time implies that the disease progresses slowly. Hence those with longer pre-clinical sojourn times are likely to live longer compared to individuals with short pre- clinical durations. Consequently, since the screening exam is likely to diagnose individuals with longer pre-clinical sojourn times, this group will tend to have longer survival regardless of treatment.

Figure 1 depicts a typical history of an individual who may be diagnosed by an early detection program. The interval between time of inception of disease and clinical diagnosis corresponds to the pre-clinical disease duration (sojourn time in the pre-clinical state).

Now consider a population of people with pre-clinical disease as depicted in Figure 2. According to our model the length of a horizontal line is the pre- clinical sojourn time. The pre-clinical sojourn time will differ among individuals and follows a probability distribution. Suppose an early detection modality has unit sensitivity. Then case finding is equivalent to placing a vertical line located at a random time point. The intersection of the vertical line with a horizontal line on the figure corresponds to a diagnosed case. Clearly the vertical line is more likely to intersect a longer horizontal line than a shorter horizontal

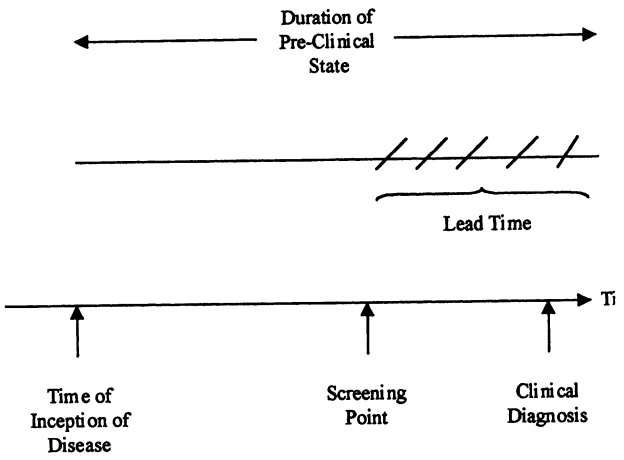
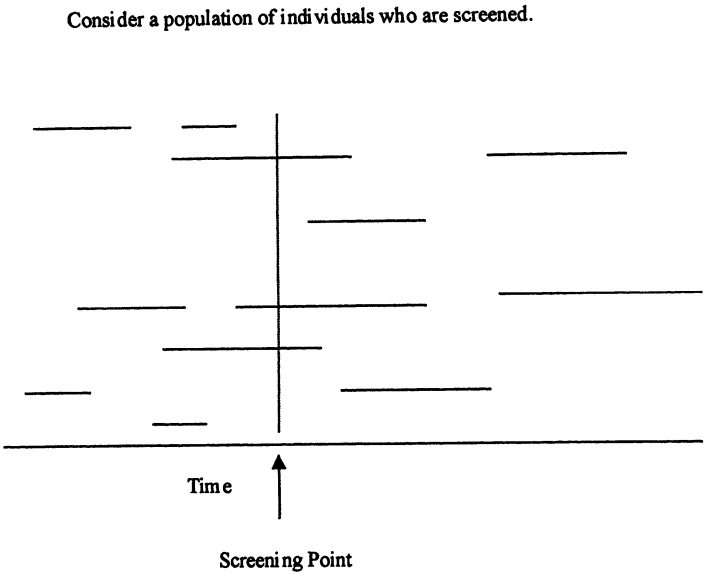


Figure 1.1. Relationship between duration of Preclinical Disease, screening point and lead time



Each horziontal line represents duration of preclinical disease for an individual.
OBSERVE: Vertical dotted line has a higher Probability of intersecting horizontal line; i.e., the screening procedure finds those individuals having longer pre-clinical durations.

Figure 1.2. Population of individuals with varying durations of pre-clinical disease

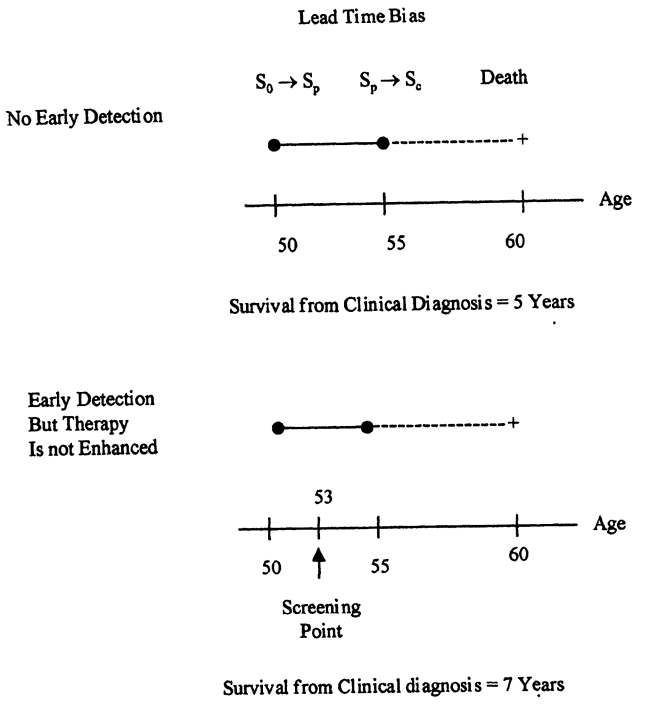


Figure 1.3. Illustration of Lead Time Bias. The top diagram illustrates the history of an individual who is diagnosed in the clinical state at age 55 and dies at age 60. The bottom diagram illustrates the natural history of the same individual who enters the pre-clinical state at age 50, diagnosed by an early detection exam at age 53, but dies at age 60 because of the occurrence of metastatic disease before age 53.

line and results in a length biased sample of diagnosed cases. The length biased phenomenon occurs irrespective of the value of the exam sensitivity.

The lead time bias is another bias associated with the comparison of survival duration. The survival duration is the time between disease diagnosis and death. If the early detection procedure does indeed find disease earlier compared to diagnosis with routine medical care, and there is no enhancement with therapy, a person will not benefit. Nevertheless because of earlier case finding, the survival will appear to be longer even though there is no benefit. Figure 3 depicts the situation for a typical case.

2.3 The Stable Disease Model

The transitions to S_p and S_c depend on time which may refer to age or chronological time. The stable disease model is defined by the transitions being independent of time. If $P(t)$ is defined as the probability of being in S_p at time t

(prevalence function), then the stable disease model is defined by $P(t)$ being independent of t, i.e., $P(t) = P$. This implies that the incidence of the disease is also independent of time. When the stable disease model is true we have the relationship

$$P = mI \tag{1.1}$$

where m is the mean sojourn time in S_p and I is the incidence of the disease.

The stable disease model does not hold for most cancers as the incidence of the disease is age dependent. However if the marginal distribution over all ages is independent of chronological time, the stable disease model will hold. That is, if $P(t)$ and $I(t)$ refer to the prevalence and incidence of disease at age t, then $P = \int_0^\infty P(t)dt$ and $I = \int_0^\infty I(t)dt$ will satisfy the relation $P = mI$. Implicit in this relationship is that the incidence of the disease is not related to chronological time.

3. LENGTH BIASED SAMPLING

Cases diagnosed by a screening exam are not diagnosed at random as discussed in 2.2, but constitute a length-biased sample. The longer the individual is in the pre-clinical state the greater the chance of being diagnosed by a scheduled exam. If the clinical course of the disease is positively correlated with the pre-clinical course, then individuals diagnosed in an early detection program are likely to live longer because of "slow" growing disease. As a result, survival is an inappropriate endpoint for evaluating the benefit of an early detection program. In this section we illustrate the main statistical properties of length biased sampling.

Length biased sampling assumes that the probability of a case being diagnosed is proportional to the sojourn time in S_p. We can write the probability of a case being diagnosed in S_p by defining T as the random variable of the sojourn time in S_p having probability density function $q(t)$ and

$$a = \{ \ 1 \quad \text{if diagnosed in } S_p 0 \text{otherwise.}$$

Hence if t is fixed, the probability of a case being diagnosed in S_p is

$$P\{a = 1 | t < T \leq t + dt\} \propto t.$$

Therefore the joint probability is

$$P\{a = 1, t < T \leq t + dt\} \propto tq(t)dt$$

resulting in

$$P\{t < T \leq t + dt | a = 1\} = tq(t)dt/m \tag{1.2}$$

where $m = \int_0^\infty tq(t)dt$. The mean of the sojourn time in S_p conditional on being diagnosed early is

$$E(T|a=1) = E(T^2)/m = m(1+C^2),\ C = \sigma/m$$

which shows that this sojourn time is larger than the sojourn time in the general population.

Two other relevant random variables are the forward and backward recurrence times. The backward recurrence time is the length of time an individual is in the pre-clinical state until the time of early diagnosis. The forward recurrence time is the time from early diagnosis to when the disease would have been routinely diagnosed in the clinical state. We shall assume that the timing of the examination is made independent of when an individual entered the pre-clinical state. Hence the diagnosis of disease could be made at any time point in the pre-clinical state. This can be modeled by assuming that the backward recurrence time follows a uniform distribution conditional on the time in the pre-clinical state. Defining T_b and T_f as the random variables for the backward and forward times we can write $P\{x < T_b \leq x + dx | t < T \leq t + dt, a = 1\} = dx/t, 0 < x \leq t$, from which the joint distribution of (T_b, T) is

$$\begin{aligned}
&P\{t < T \leq t + dt, x < T_b \leq x + dx | a = 1\} \\
&= P\{t < T \leq t + dt | a = 1\} \cdot P\{x < T_b \leq x + dx | t < T \leq t + dt, a = 1\} \\
&= \frac{tq(t)}{m} dt \cdot \frac{dx}{t},\ 0 < x \leq t < \infty \\
&= \frac{q(t)}{m} dt dx,\ 0 < x \leq t < \infty. \qquad (1.3)
\end{aligned}$$

Thus the pdf of T_b is

$$q_b(x) = \int_x^\infty \frac{q(t)}{m} dt = Q(x)/m \qquad (1.4)$$

having the expected value

$$E(T_b) = \int_0^\infty \frac{xQ(x)}{m} dx = \frac{m}{2}(1+C^2), C = \sigma/m. \qquad (1.5)$$

Since $T_b + T_f = T$, the distribution of the forward recurrence time distribution is the same as the backwards recurrence time distribution. The joint distribution of the forward and backward recurrence times is an immediate consequence of (3); i.e.,

$$P\{x < T_b \leq x + dx, y < T_f \leq y + dy | a = 1\} = \frac{q(x+y)}{m} dx dy \qquad (1.6)$$

In general, the backward and forward recurrence times are not independent. However when $q(t) = \lambda e^{-\lambda t}$ we have

$$\frac{q(x+y)}{m} = \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda y}, \quad m = 1/\lambda$$

and thus they are independent.

The relations derived above enable the estimation of the mean of the forward recurrence time distribution by using the mean ages of diagnosis. Assume that there are two comparable groups — one of which is a control group of individuals receiving routine care and the other is a group receiving a single early detection examination. This could arise in a clinical trial in which two groups are randomized to either a control or an early detection group. Let the mean time in the pre-clinical state be designated by m and define m_c and m_e to be the mean ages of diagnosis for the control and the early detection groups. If A is the mean age at which individuals enter the pre-clinical state, then

$$m_c = A + m \quad , \quad m_e = A + m_b$$

where m_b is the mean backward recurrence time. Since our derivation showed that $m_b = m_f$ we have an expression for m_f in terms of quantities which may be estimated; i.e.

$$m_f = m - (m_c - m_c). \tag{1.7}$$

It can also be shown that the necessary and sufficient condition for the sojourn time to follow an exponential distribution is that $m_c = m_e$ in which case $m_f = m$. Also the bounds on m_f are $m/2 \leq m_f \leq m$.

Data on the average age of diagnosis is available from the clinical trial carried out by the Health Insurance Plan of New York to compare the benefit of a combined physical and mammogram examination versus usual medical care Shapiro et al. (1977, 1988). In the first year of the study the average age of those detected by the initial exam was $\hat{m}_e$=53.8 years $(n = 54)$ versus $\hat{m}_c = 53.3$ years $(n = 45)$ for those in the control group; cf. Zelen and Feinleib (1969). Clearly these ages are comparable and it is possible to conclude that the sojourn time in the pre-clinical state can be assumed to follow the exponential distribution.

The importance of the forward recurrence time distribution arises because it is the probability distribution of the time gained by early diagnosis. The mean forward recurrence time is an indirect measure of the efficacy of a screening program. The forward recurrence time cannot be observed directly as once a disease is diagnosed its natural history is changed by treatment. However equation (7) offers a way to estimate the mean forward recurrence time. Suppose a randomized clinical trial to evaluate an early detection program has a control

group and a study group. At the first examination, the average age of detection (m_e) can be estimated and after some period of time the average age of diagnosis (m_c) can be estimated from the control group. It remains to determine m. If the sensitivity of the exam is unity and the model is approximated by the stable disease model, then m may be estimated by P/I where P is the probability of being diagnosed in the special exam and I is the disease incidence. If the sensitivity is not unity, then P/I estimates βm where β is the exam sensitivity. Estimating the mean is a more complicated task. Section 6 discusses methods for estimating β and m.

4. PLANNING CLINICAL TRIALS TO EVALUATE EARLY DETECTION PROGRAMS

4.1 Background

Randomized clinical trials are the principal way to evaluate the benefits of an early detection program. A typical clinical trial consists of randomizing individuals without disease to two groups. One group is a control group receiving usual medical care; the other is a study or screened group receiving one or more periodic exams. The endpoint is death and the mortalities of the two groups are compared. One of the principal reasons for the necessity of randomized trials is that length and lead time biases influence the comparisons in non-randomized trials and cannot be eliminated. As mentioned earlier, survival is not an appropriate endpoint for group comparisons. However mortality comparisons eliminate the length and lead time biases.

Randomization may be carried out in which the individual is the unit being randomized or by group randomization. Trials which have been carried out in Sweden on breast cancer have utilized group randomization where the unit is a county or a designated region. Individuals residing in regions randomized to periodic screening will receive invitations to participate in a screening program. The remaining regions serve as a contorl group when individuals have their usual medical care. After a period of time the mortality of the two sets of regions are compared. These comparisons depend on the quality of the vital statistics system to track all residents in the regions who are eventually diagnosed with the disease under study. Clearly compliance is a major issue with the group randomization trials as many residents in the study group may ignore the invitation to participate in an early detection program. Some questions also arise about the consent process for group randomized studies. Presumably those individuals volunteering to undergo early detection exams have implicitly given consent. However members of the control group are not approached for consent. This system appears to violate the Code of Federal Regulations governing all human experimentation in the United States. The code requires

that all individuals participating in a study must give consent. Randomizing by group is not recommended as it violates guidelines on consent.

Nearly all of the randomized early detection trials have been planned using ideas from the planning of therapeutic trials. In therapeutic trials: (i) all individuals have disease, (ii) longer follow-up time results in greater power and (iii) equal size groups result in maximum power. These features are not present or are incorrect in early detection trials. Early detection trials are characterized by (i) all individuals should not have disease, (ii) long follow- up time may result in a reduction of power and (iii) equal size groups do not result in maximum power for fixed costs.

Moreover early detection trials have features which must be taken into account in the planning of the trials. Some of these features are: (i) the necessity of choosing the number of exams and the spacing between exams in the study group; (ii) calculation of the time to analyze the study to achieve maximum power (analyses times which are too short or long will have reduced power); (iii) sensitivity and specificity of the exams, prevalence and incidence of the clinical trial population; and (iv) the sample sizes for each group.

4.2 Examples

We shall illustrate the impact of some of these factors by two examples. For purposes of calculation, it will be assumed that all distributions (sojourn time and survival) are exponential. It will be assumed that the median survival time for those in S_p at the initial exam is different than those who enter S_p after the trial started. For those in S_p, at the time of the first exam, the median survivals are 11 years (control) and 20 years for those diagnosed in the study group by exam. Those individuals entering S_p after the initial exam have median survivals of 10 (control) and 17 years (diagnosed by exam). The stable disease model will be used with the prevalence $P = 0.007$ and $w(t) = .002$ giving rise to a median sojourn time of 2.4 years. These parameters have been chosen to reflect parameter values reported in the breast cancer early detection literature; cf. Walter and Day (1983), Shapiro et al. (1977) and Stomper and Gelman (1989).

The basic paper which formulates the problem of planning early detection clinical trials is by Hu and Zelen (1997). The two examples in this section are taken from that paper. In all that follows it will be assumed that the trials have been planned to have a one-sided type I error of 5% and that the statistical test compares the proportion of deaths due to the particular disease being screened. Also the typical randomized clinical trial consists of two groups which are referred to as a control group (usual medical care) and a study group (receiving one or more periodic examinations). Accrual in the trial will be uniform over a three-year period.

Example 1: Equal numbers in study and control groups. The first example will illustrate how power changes as sample size, number of periodic examinations and time between periodic examinations are varied. We consider equal sample sizes for both control and study groups. All follow-up times are dated from the end of patient accrual.

For a fixed sample size of 25,000 in each group, Figure 4(a), (b) shows the power for two and four periodic examinations respectively with different fixed intervals between examinations and varying years of follow-up. In Figure 4(a), as the time interval between examinations increase, the optimal power, top curve, increase as well as the required optimal years of follow-up. If the follow-up time is not optimal, other two curves, the power increases until a maximum is reached for a specific time between examinations. Increasing the spacing between examinations results in decreased power. Figure 4(b) has similar interpretations. However, for the four-examination study, at 5 and 10 years follow-up, with the time interval between examinations being 2-5 years, not all of the participants receive the indicated number of examinations. The fractions in the figures indicate the proportion of participants receiving all scheduled examinations. When the sample size and follow-up time are fixed, these figures can be used to determine the combination of the number of examinations and the time interval between examinations resulting in maximum power. For example, if the sample size is 25,000 and follow-up time is 10 years, the combination of four examinations, spaced two years apart, achieves the maximum power of 80%: close scrutiny of Figure 4(b) shows that the curve for a follow-up time of 15 years has slightly less power for cases for which the optimal follow-up time is smaller.

Figure 5(a), (b) shows the power for sample sizes of 25,000, 50,000, and 75,000 in each group, for different fixed intervals between examinations and varying years of follow-up, when three examinations are offered in the study group. As the follow-up time increases the pattern of the power curve changes. For five years of follow-up, the power decreases as the time interval between examinations increases. For fifteen years of follow-up, the power increases as the time interval between examinations increases and eventually decreases as the time interval becomes very long. These figures illustrate how to design trials to achieve a pre-determined power. For example, if the required power is at least 90%, with a five-year follow-up period and the examination intervals being one or two years apart, then this power will be attained for a sample size of 75,000. Similarly, when the follow-up time is fifteen years and sample size is 50,000 (75,000), the power of at least 90% will be attained for spacing between examinations of 2-5 (1-5) years. However this level of power is never attained when the sample size is 25,000.

All the figures illustrate the trade-offs between the various factors associated with the planning of early detection trials. Now suppose the aim is to achieve a

Table 1.1. Example 1: Combination of factors resulting in power of 80% and 95% for one-sided 0.05 level of significance, with equal sample sizes in the study and control groups

		Power of 80%		Power of 95%	
Sample Size	No. of exams	Time between exams (yrs)	Optimal follow-up time (yrs)	Time between exams (yrs)	Optimal follow-up time (yrs)
25,000	4	2	10	-	-
50,000	2	2	7-8	-	-
	3	1	5-6	3-4	12-13
	4	$\frac{1}{2}$	6	2	8-9
75,000	2	$\frac{1}{2}$	6	3-6	10-13
	3	$\frac{1}{2}$*	4	1	7
	4	$\frac{1}{2}$*	4	$\frac{1}{2}$	8

power of 95% with a 0.05 one-sided level of significance. Table 1 summarizes the trials which will achieve this goal. It also contains designs if a power of 80% is required.

* Power $> 80\%$

In general, the larger the interval between examinations, the longer is the required follow-up time to achieve optimal power. Also the smaller the sample size, the more critical is the choice of the number of examinations and the time between examinations.

Example 2: Unequal numbers in study and control groups. Our next example illustrates possible gains that can be made by not having equal numbers of participants for the study and control groups. It will be assumed that the principal costs for the study groups are the number of examinations in the study group. Hence the principal costs associated with a study group will be the same whether 50,000 people receive only one examination, 25,000 people receive two examinations, etc. Thus if n_k denotes the number of people receiving k examinations ($k = 1, 2, \ldots$), the principal costs will be the same, provided $kn_k = n_1$.

In this example we will illustrate the optimal choices for k when $n_1 = 50,000$. This sample size also refers to the number of people in the control group. Table 2 summarizes the calculations for $k = 1(1)4$ examinations with $n_k = n_1/k$. It is clear that the power increases as k increases. Also an increase in the number of examinations results in larger follow-up times, in almost all cases, in order to achieve maximum power.

Suppose now a study is required which has 80% power at a 0.05 one-sided level of significance. Table 3 summarizes all combinations of factors and study group costs which achieve this power. The entry under the heading "cost of study group" is the product kn_k. The table shows that the lowest cost experimental