Scheduling Periodic Examinations for the Early Detection of Disease: Applications to Breast Cancer

Sandra J. LEE and Marvin ZELEN

This article develops and extends previous investigations on stochastic models for selecting examination schedules targeted at earlier diagnosis of chronic diseases. The general aim is to provide guidelines for public health programs in the choice of examination schedules. The main features of such schedules are the initial age to begin a scheduled examination program, the intervals between subsequent examinations, and the number of examinations. Our basic model consists of three health states: a disease-free or nondetectable state; a preclinical state, in which an individual has disease but is asymptomatic and is unaware of it; and a clinical state in which the disease has been diagnosed by routine methods. The aim of early detection programs is to identify individuals in a preclinical local state, which may result in higher cure rates or longer survival. We introduce two basic ideas that either individually or together can lead to satisfactory examinations schedules. The *threshold method* constructs examination schedules so that the probability of an individual being in the preclinical state is always bounded by a preselected value. The concept of *schedule sensitivity* is the ratio of the expected number of cases diagnosed on scheduled examinations to the expectation of the total number of cases. Combining the threshold and schedule sensitivity methods enables an assessment of the trade-offs with regard to the initial age at examination and cost per case found. All of these methods are extended to high-risk populations. The schedule sensitivity, cost per case diagnosed on a scheduled exam, and the total cost can be calculated for any early detection program. This enables comparisons of different strategies. We illustrate the applicability of our methods to scheduling examinations for female breast cancer.

KEY WORDS: Early detection programs; Periodic screening.

1. INTRODUCTION

The main motivation of screening for chronic diseases is to diagnose disease early in the expectation that benefit from earlier treatment will be enhanced, resulting in reduced mortality. With advances in diagnostic testing techniques, early diagnosis of breast, ovarian, cervical, colorectal, and other cancers has become more feasible. For example, screening asymptomatic individuals for breast cancer by mammography and/or clinical breast examination improves the chance of detecting breast cancer at an earlier stage, before metastases occur. The evidence for decreased breast cancer mortality following early detection has been demonstrated for women aged 50 years and older in a series of early detection randomized clinical trials (for a summary see Fletcher, Black, Harris, Rimer, and Shapiro 1993).

Planning public health programs for the early detection of a chronic disease requires determining (a) the ages (or combination of risk factors) which indicate when individuals should begin to participate in an early detection program, (b) the spacings between examinations and their relation to age, and (c) recommendations for screening schedules for high-risk individuals. Our view is that any public health program for the early detection of disease represents a health policy decision. In general, a public health screening program should be planned so as to be "optimal" with regard to some utility function. Determining "optimal screening programs" utilizing clinical trials or other clinical studies is not feasible as the number of potential combinations relating to age, examination spacings, and risk factors is very large. Such studies would require long follow-up times and very high costs. The only reasonable approach to this problem is to use analytical models of the early detection process that make use of whatever meaningful biological data are available.

Many investigators have studied mathematical models of the screening process, including Albert, Gertman, Louis, and Liu (1978), Baker and Chu (1990), Day and Walter (1984), Dubin (1981), Eddy (1980, 1983), Eddy and Shwartz (1982), Habbema, VanOortmarssen, Lubbe, and van der Mass (1984), Kirch and Klein (1974), Knox (1973), Parmigiani (1993, 1997), Prorok (1976a,b), Shwartz (1978), Shwartz and Plough (1984), Zelen (1993), and Zelen and Feinleib (1969). However, with the exception of Baker and Chu (1990), Parmigiani (1993, 1997), and Zelen (1993), these authors have paid little attention to optimizing examination schedules.

Another aspect to this problem is that the earlier the detection of disease, the more likely the individual will be in a curative disease state. Many screening programs undoubtedly detect disease early, which may result in a higher proportion of cases with localized disease. For example, in breast cancer, it has been demonstrated that a higher proportion of localized cases (~80%) are detected in a scheduled screening exam than through usual medical care (~50%). In this article we assume that evidence (mainly from clinical trials) has already been generated showing significant benefit for early diagnoses of disease. The next step is to extend this benefit to the population at large with early detection public health programs. The goal of such a program is to diagnose as many individuals as possible at scheduled examinations. We propose methods to guide public health

Sandra J. Lee is Research Scientist (E-mail: sjlee@jimmy.harvard.edu) and Marvin Zelen is Professor of Statistical Science, Department of Biostatistics, Harvard School of Public Health and Dana Farber Cancer Institute, Boston, MA 02115. The authors acknowledge discussions with Rebecca Gelman and Steven Skates and comments by Stuart Baker that have clarified many of the issues. This work was partially supported by research grants from the National Cancer Institute and the Massachusetts Department of Public Health.

Journal of the American Statistical Association
December 1998, Vol. 93, No. 444, Applications and Case Studies

policy in comparing different early detection programs. Our methods are motivated from cancer—specifically, the early detection of breast cancer using mammograms. However, the ideas are general and can be applied to all chronic diseases.

We also address the issue of screening for high-risk groups. It is possible to identify subpopulations who are at higher lifetime risk for many chronic diseases. In the past, such subpopulations have been identified when the disease has a familial component. However, recently, several genetic markers have been found that are specific for some chronic diseases. Such findings are likely to increase in the future. This raises the problem of recommending examination schedules for these high-risk populations. Currently no theory deals with this problem. Earlier diagnosis of these high-risk individuals, when they are in an early disease stage, may result in improved prognosis. In this article we also apply our modeling considerations to these high-risk populations.

The article is organized as follows. In Section 2 we develop a statistical model for the screening process and also develop the threshold and schedule sensitivity methods. We illustrate our model for schedules targeted at the early diagnosis of breast cancer in Section 3. Calculations show how schedules change under various parameter assumptions. Schedules for high-risk groups are considered, and the effectiveness, and robustness of these schedules are evaluated. We propose an analytical approach for planning screening programs in Section 4, and conclude with a discussion in Section 5.

2. MODEL

2.1 Notation

Our development of the model for screening closely follows the work of Zelen (1993). Consider that an individual's health may be in one of three states; the undetectable or disease-free state S_0, the preclinical state S_p, and the clinical state S_c. S_p refers to an asymptomatic and detectable disease state; S_c refers to a symptomatic disease state in which the disease is diagnosed by routine methods. We assume that the disease progresses in the manner of $S_0 \to S_p \to S_c$. The goal of screening programs is to detect disease in the preclinical state S_p. The sensitivity of the diagnostic test used in screening is denoted by β; that is, β denotes the probability of a screening examination finding an individual in S_p conditional on being in S_p. Let $P(t)$ denote the probability of being in S_p at age t. Also, define $w(t)\Delta t$ and $I(t)\Delta t$ as the transition probabilities, $S_0 \to S_p$ and $S_p \to S_c$, during an age interval $(t, t+\Delta t)$. Let $q(z)$ be the probability density function of the sojourn time in S_p and let $Q(z) = \int_z^\infty q(x)\,dx$. The time variable t can refer to chronological time or age; here, it refers to age. We assume that $(n+1)$ screening examinations will be carried out at ages $t_0 < t_1 < \cdots < t_n$. The time t_0 refers to an individual's age at the initial screening exam. We define the ith interval as $(t_{i-1}, t_i]$ and an ith generation individual as someone who enters S_p during this interval for $i = 1, 2, \ldots, n$. The 0th generation includes all the individuals who enter S_p before t_0. We define $t_{-1} = 0$ and the 0th interval by $(0, t_0]$.

2.2 Basic Model

We define the forward recurrence time for an ith generation individual as the time in S_p taking t_i as the origin or, equivalently, the time in S_p excluding time spent in the ith interval. For the ith generation, let $q_i(z), i = 0, 1, \ldots, n$, denote the probability density function of the forward recurrence time and let $P_i(t_i)$ denote the probability of an ith generation individual being in S_p at age t_i. Then the joint probability of an ith generation individual being in S_p at time t_i and having a forward recurrence time z is

$$P_i(t_i)q_i(z) = \int_{t_{i-1}}^{t_i} w(x)q(t_i - x + z)\,dx \qquad (i = 0, 1, \ldots, n),$$

where

$$P_i(t_i) = \int_{t_{i-1}}^{t_i} w(x)Q(t_i - x)\,dx.$$

The joint probability of an ith generation individual being in S_p at age t_i and having a forward recurrence time of at least $(t - t_i)$ is given by

$$P_i(t_i)Q_i(t - t_i) = \int_{t_{i-1}}^{t_i} w(x)Q(t - x)\,dx.$$

For a given diagnostic test with sensitivity β, we are interested in detecting individuals in S_p. At age t, with $t_{r-1} \le t < t_r$, the probability of having undiagnosed preclinical disease after having r examination is given by

$$\begin{aligned} P(t|r) &= (1-\beta)^r P_0(t_0)Q_0(t - t_0) + \Phi(r-2) \\ &\quad \times \sum_{k=1}^{r-1} (1-\beta)^{r-k} P_k(t_k)Q_k(t - t_k) \\ &\quad + \int_{t_{r-1}}^{t} w(x)Q(t - x)\,dx, \end{aligned}$$

which can be written as

$$\begin{aligned} P(t|r) &= (1-\beta)^r P_0(t_0)Q_0(t - t_0) + \Phi(r-2) \\ &\quad \times \sum_{k=1}^{r-1} (1-\beta)^{r-k} \int_{t_{k-1}}^{t_k} w(x)Q(t - x)\,dx \\ &\quad + \int_{t_{r-1}}^{t} w(x)Q(t - x)\,dx, \end{aligned} \tag{1}$$

where

$$\Phi(x) = \begin{cases} 1 & \text{if } x \ge 0 \\ 0 & \text{otherwise.} \end{cases}$$

The quantity (1) captures the probability of the kth generation individual not being detected in the first $(r-k)$ examinations and having a forward recurrence time of at least $(t - t_k)$ for $k = 0, 1, \ldots, r-1$, and the rth generation individual entering S_p during (t_{r-1}, t). Finally, we sum the contributions over each generation. Note that for $t_n \le t < \infty$, we have the expression $P(t|n+1)$ with $r = n+1$, which gives the probability of having disease after the last scheduled examination.

Finally, defining $S(t)$ as the probability of having undiagnosed preclinical disease at age $t > t_0$, we can use (1) to obtain

$$S(t) = \sum_{r=1}^{n} \Phi(t - t_{r-1}) \times \{1 - \Phi(t - t_r)\}P(t|r) + \Phi(t - t_n)P(t|n+1).$$

The derivation of (1) assumes that the detection process for an individual consists of Bernoulli trials. Realistically, the sensitivity may have a distribution centered around a value β. As an example, consider a women who undergoes periodic mammographic examinations to detect breast cancer. Every examination results in a fresh mammogram. If the woman has a lesion, then positioning of the breast and the lesion may be different for every mammogram. Many studies have demonstrated that lesions may not be detected in a mammogram but can be detected by a physical exam and vice versa. This can be ascribed to the positioning of the lesion. The situation is further complicated because it is well known that having more film readers increases the sensitivity. Furthermore, the more experienced the reader, the higher the sensitivity. The current technology of mammograms is that very small lesions can be detected once the size reaches a threshold value. The diagnosis depends on positioning, the spatial orientation of the tumor in the breast, and the experience of the reader. Although increasing tumor size may theoretically raise the probability of detection, nevertheless it is the positioning and experience of the reader that dominate the sensitivity of the exam process after the tumor has reached a threshold value. Because periodic examinations involve new mammograms, and for all practical purposes different readers are chosen at random, the assumption of a sequence of Bernoulli trials as a model for periodic examinations is a reasonable approximation to the observed phenomenon.

2.3 Calculation of $w(t)$

Calculation of $S(t)$ requires knowledge of $w(t)$, but one cannot observe $w(t)$ directly. However, data are available on age-specific incidence ($S_p \to S_c$), which enables the estimation of $w(t)$. We show how knowledge of the age-specific disease incidence and an assumed form of the sojourn time distribution allows calculation of $w(t)$. Define $I(t)\,dt$ to be the probability of transition from S_p to S_c during the age interval $(t, t+dt)$. Then the equation connecting $I(t), w(t)$, and $q(z)$ is

$$I(t) = \int_0^t w(z)q(t-z)\,dz.$$

Usually the incidence data are grouped by age. Let the jth age interval be defined as $A_j = (x_{j-1}, x_j]$ for $j = 1, 2, \ldots, k$, where $x_0 = 0$. Throughout we assume constant intervals and that $\Delta = (x_j - x_{j-1})$. The incidence of disease for the jth age group is then defined by

$$I_j = \int_{A_j} I(t)\,dt.$$

Although $w(t)$ may be a continuous function of age, the available grouped data on incidence only allow $w(t)$ to be estimated as a step function. For this purpose, define $w(t) = w_j$ for $t \in A_j$.

If $t \in A_j$, then

$$I(t) = \int_0^t w(z)q(t-z)\,dz = w_1\{1 - Q(t)\} \quad \text{for} \quad j = 1$$

and

$$\begin{aligned} I(t) &= \sum_{i=1}^{j-1} \int_{A_i} w(z)q(t-z)\,dz \\ &\quad + \int_{x_{j-1}}^{t} w(z)q(t-z)\,dz \quad \text{for} \quad j > 1 \\ &= \sum_{i=1}^{j-1} w_i\{Q(t-x_i) - Q(t-x_{i-1})\} \\ &\quad + w_j\{1 - Q(t-x_{j-1})\}. \end{aligned}$$

Hence on integrating over A_j, we have

$$I_1 = \int_{A_1} I(t)\,dt = w_1 m\left[\frac{\Delta}{m} - \{1 - Q_f(\Delta)\}\right] \quad \text{for} \quad j = 1$$

and

$$\begin{aligned} I_j &= \int_{A_j} I(t)\,dt \quad \text{for} \quad j > 1 \\ &= m\sum_{i=1}^{j-1} w_i[Q_f\{(j-i+1)\Delta\} \\ &\quad + Q_f\{(j-i-1)\Delta\} - 2Q_f\{(j-i)\Delta\}] \\ &\quad + w_j m\left[\frac{\Delta}{m} - \{1 - Q_f(\Delta)\}\right], \end{aligned} \tag{2}$$

where $Q_f(t) = \int_t^\infty Q(x)/m\,dx$ and $m = \int_0^\infty Q(x)\,dx$. The reason for adopting the notation $Q_f(t)$ is that under some circumstances, the forward recurrence pdf of the sojourn time distribution in S_p for the 0th generation is $q_f(t) = Q(t)/m$. Here $Q_f(t)$ is the tail area of the forward recurrence time distribution. Note that if $q(t) = \lambda\exp(-\lambda t)$ $(t \geq 0)$, then $Q_f(t) = \exp(-\lambda t)$, where $m = 1/\lambda$. The expression

$$Q_f\{(j-i+1)\Delta\} + Q_f\{(j-i-1)\Delta\} - 2Q_f\{(j-i)\Delta\} \tag{3}$$

in (2) is the second difference of $Q_f(t)$ where $t = (j-i+1)\Delta$. Thus if the sojourn time distribution is exponential, then (3) can be written as $\exp\{-\lambda(j-i)\Delta\}\{\exp(-\lambda\Delta) + \exp(\lambda\Delta) - 2\}$. Equation (2) allows the quantity w_j to be calculated recursively if the sojourn time distribution is known.

2.4 Threshold Method for Scheduling Examinations

Consider the age interval $[t_0, T_0]$, which is termed the screening horizon. This is the interval in which examinations are to be scheduled. The threshold method involves

choosing a threshold value (P_0) and ages of examination $\{t_i\}$ such that $S(t_i) = P_0$ for $t_0 \le t_i \le T_0$. One possible threshold value is the probability of being in the preclinical state at the time of the initial examination; that is, $P_0 = P_0(t_0)$. This strategy is suitable when the probability of being in the preclinical state increases with age. Then the examination schedule guarantees that within the age interval $[t_0, T_0]$, $S(t) \le P_0(t_0)$. Hence if t_i corresponds to the age at the $(i+1)$th examination, then we will choose t_i so that $S(t_i) = P_0(t_0)$. Note that the function $S(t)$ is such that $S(t_r) > S(t_{r^+})$; that is, after every exam the function decreases and then increases as t increases. When it reaches $P_0(t_0)$, another exam is given, and the process repeats itself.

High-risk individuals are characterized by an increased incidence relative to the general population. If $I^*(t)$ corresponds to the incidence for high-risk individuals and $P_0(t_0)$ is the threshold value (for individuals not at high risk), then we can adapt the threshold method to find examination schedules for these high-risk individuals. Ordinarily, one may not know the age-specific incidence for high-risk individuals but may have an estimate of the increased lifetime risk of the disease. If $\int_0^\infty I^*(t)\,dt = \rho \int_0^\infty I(t)\,dt$, then a reasonable approximation to $I^*(t)$ is that $I^*(t) = \rho I(t)$, where ρ represents the multiple for the increased risk. Hence replacing I_j by ρI_j in (2) enables calculation of $w_i^* = \rho w_i$ for high-risk individuals. The corresponding examination schedule will have an initial examination at age t_0^*, where at this age, the probability of a high-risk individual being in S_p is $P_0(t_0^*) = P_0(t_0)$. Of course, $P_0(t_0^*)$ can be set at any threshold level and need not be $P_0(t_0)$.

2.5 Effectiveness of a Screening Program (Schedule Sensitivity)

There are many measures of the effectiveness of early detection public health programs. Possible measures or utility functions include proportion of cures, years of life saved, and number of subjects found with disease at a scheduled examination. Adopting some of these utilities may require information on cure rates or survival both with and without special examinations. Such data may not be available, or if available may have significant biases so as to throw doubt on any utility optimization using such data. Furthermore, different utility functions may result in radically different early detection programs. For example, if the utility function were to diagnose large numbers of individuals by special examinations, then one would target older people if disease incidence increases with age. Alternatively, if the utility function were years of life saved, then younger people would be targeted in an early detection program.

The most basic information available for planning early detection programs is age-specific disease incidence, sensitivity of an examination, and mean sojourn time in the preclinical state. Our development uses these parameters. Age-specific disease incidence is routinely available for nearly all chronic diseases; data on the sensitivity of an examination and the mean sojourn time in the preclinical state (as well as the parametric form) may be obtained from well-conducted early detection trials.

One measure of the effectiveness of a screening program is the ratio of the expected number of cases diagnosed at a scheduled examination relative to the expected number for all cases for a specified screening horizon. This is referred to as the *screening horizon schedule sensitivity.* To calculate the screening horizon schedule sensitivity, both the expected number of interval cases and cases found on an early detection examination are needed. Let $I_r(t)dt$ be the probability that an individual is incident within the interval $(t, t+dt)$, where $t_{r-1} \le t < t_r$. Then the total probability of being an interval case in the rth interval is $I_r = \int_{t_{r-1}}^{t_r} I_r(t)\,dt$, and the total probability of being an interval case over the interval (t_0, t_n) is $I = \sum_{r=1}^{n} I_r$. If $t_n = T_0$, where T_0 is the upper endpoint of the screening horizon, then no further considerations are required. However, if $t_n < T_0$, then the probability of being incident in the interval $(t_n, T_0]$ must be computed.

To find $I_r(t)$, we note that an ith generation individual who is incident in the rth interval $(i \le r \le n)$ must have failed to be detected at $(r-i)$ previous exams and have a forward recurrence time $(t - t_i)$. In addition, individuals can enter S_p in the rth interval and be incident in that interval. Hence we can write

$$I_r(t) = \sum_{i=0}^{r-1} (1-\beta)^{r-i} P_i(t_i) q_i(t - t_i) + \int_{t_{r-1}}^{t} w(x) q(t-x)\,dx \qquad (r = 1, 2, \ldots, n)$$

and

$$\begin{aligned} I_r &= \int_{t_{r-1}}^{t_r} I_r(t)\,dt \\ &= \sum_{i=0}^{r-1} (1-\beta)^{r-i} P_i(t_i)\{Q_i(t_{r-1} - t_i) - Q_i(t_r - t_i)\} \\ &\quad + \int_{t_{r-1}}^{t_r} w(x)\,dx - P_r(t_r). \end{aligned}$$

If $t_n < T_0$, then we must also calculate the probability of incidence in the interval $(t_n, T_0]$. This interval is designated as the $(n+1)$th interval. Using similar reasoning as in deriving $I_r(t)$, we have

$$I_{n+1}(t) = \sum_{i=0}^{n} (1-\beta)^{n+1-i} P_i(t_i) q_i(t - t_i) + \int_{t_n}^{t} w(x) q(t-x)\,dx$$

for $t_n \le t < T_0$ and

$$\begin{aligned} I_{n+1} &= \int_{t_n}^{T_0} I_{n+1}(t)\,dt \\ &= \sum_{i=0}^{n} (1-\beta)^{n+1-i} P_i(t_i)\{Q_i(t_n - t_i) - Q_i(T_0 - t_0)\} \end{aligned}$$

$$+ \int_{t_n}^{T_0} w(x)\,dx - \int_{t_n}^{T_0} w(x)Q(T_0 - x)\,dx.$$

Denote $D_r(\beta)$ to be the probability of detection at the examination performed at age t_r. Then, using the same reasoning as before,

$$D_r(\beta) = \beta\left\{\sum_{i=0}^{r-1}(1-\beta)^{r-i}P_i(t_i)Q_i(t_r - t_i) + \int_{t_{r-1}}^{t_r} w(x)Q(t_r - x)\,dx\right\} \qquad (r = 1, 2, \ldots, n).$$

We define the screening horizon schedule sensitivity by

$$R = \frac{\sum_{r=0}^{n} D_r(\beta)}{\sum_{r=0}^{n} D_r(\beta) + \sum_{r=1}^{n+1} I_r}. \tag{4}$$

This represents the fraction of expected number of cases diagnosed on the scheduled exams relative to the total expected number of cases diagnosed during the interval $[t_0, T_0]$. In the special case where the threshold method is used with $(n+1)$ exams and the threshold probability is $P_0(t_0)$, the quantity (4) can be simplified as

$$R = \frac{(n+1)\beta P_0(t_0)}{(n+1)\beta P_0(t_0) + \sum_{r=1}^{n+1} I_r}.$$

Clearly, R is a function of the number of the exams, sensitivity, screening horizon, sojourn distribution in S_p, and age-specific incidence of disease. These ratios refer only to cases diagnosed or incident in the screening horizon $[t_0, T_0]$. However, there is also interest in having the denominator include all cases that are incident in the interval $(0, t_0)$. Denoting this incidence probability by I_0 and using the same methods as earlier, we can write

$$I_0 = \int_0^{t_0} I(x)\,dx = \int_0^{t_0} w(x)\,dx - P_0(t_0). \tag{5}$$

Thus, by summing the incidence from $r = 0$ to $(n+1)$, the *lifetime schedule sensitivity* is defined by

$$R_0 = \frac{\sum_{r=0}^{n} D_r(\beta)}{\sum_{r=0}^{n} D_r(\beta) + \sum_{r=0}^{n+1} I_r}. \tag{6}$$

The quantity (6) represents the fraction of cases diagnosed at scheduled exams during an individual's lifetime. It should be noted that the screening sensitivity can be defined over any interval and need not be confined to the screening horizon or lifetime.

2.6 Cost per Case Diagnosed

Another characterization of the diagnostic process is the cost per case found. We envision the diagnostic process as consisting of two steps. The first step is identification of subjects who test positive on the special screening exam. These subjects are then given a definitive diagnostic procedure, which is necessary to make a confirmed disease diagnosis. For example, in the case of cancerous tumors, a biopsy is ordinarily made for a definitive diagnosis.

To model these costs, define k_s as the unit cost of the special examination, k_d as the unit cost of the definitive examination, $P(+|-)$ as probability of a false positive outcome on the special exam, and $D = \sum_{r=0}^{n} D_r$. Then the *additional* cost per expected case for low-incident diseases is

$$\begin{aligned} c &= \{(n+1)k_s + P(+|-)k_d\}/D \\ &= k_d\{(n+1)k_s/k_d + P(+|-)\}/D. \end{aligned}$$

Here we have ignored the facts that cases diagnosed at a scheduled exam do not receive further exams, and that an individual may have more than one false-positive outcome in an examination program. Hence if comparisons are being made between two screening programs, denoted by superscript (1) and (2), then the relative cost per case found is

$$\frac{c_1}{c_2} = \frac{\{(n_1+1)(k_s/k_d) + P^{(1)}(+|-)\}D^{(2)}}{\{(n_2+1)(k_s/k_d) + P^{(2)}(+|-)\}D^{(1)}}. \tag{7}$$

For simplicity, the relative cost per case found can be considered without incorporating costs for false positive cases. Is this case, (7) simplifies to

$$\frac{c_1}{c_2} = \frac{(n_1+1)D^{(2)}}{(n_2+1)D^{(1)}}. \tag{8}$$

3. APPLICATIONS TO BREAST CANCER SCREENING

In this section we apply our theory to schedules for the early diagnosis of breast cancer using mammography possibly combined with physical examinations. Presently there is general agreement that mammography is beneficial in prolonging life for women aged 50 years and older, but disagreement on whether it benefits women under 50 years (National Institutes of Health Consensus Statement 1997). We limit our application to women aged 50 years and older.

3.1 Preliminaries

The first step in applying our theory is to estimate the transition probabilities $w(t)$. To estimate $w(t)$, we used the age-specific breast cancer incidence rate $I_j, j = 1, 2, \ldots, 18$, from the Surveillance, Epidemiology and End Results (SEER) program of the National Cancer Institute, 1988–1992 (1994). These incident rates are summarized in age intervals of 5 years; that is, ages 0–4, 5–9, ..., 80–84, and 85+. Using the relationship (2) we estimated w_j under the assumption of an exponential distribution for the sojourn time in S_p. There are two justifications for the choice of an exponential distribution. Zelen and Feinleib (1969) proved that the necessary and sufficient condition for the sojourn time in S_p to be the exponential distribution is that the mean age of detection in the first examination be the same as the age of detection determined by routine medical care. This condition was verified by the randomized clinical trial for breast cancer conducted by Health Insurance Plan of Greater New York (Shapiro, Venet, Strax, Venet, and Rosener 1982). Furthermore, Day and Walter (1984) derived the likelihood associated with this study and showed

Table 1. Age-Specific Breast Cancer Incidence Rates (I_j) and Transition Probabilities ($w(t)$) for $m = 4$

j	Age (A_j)	I_j for A_j	$w(t)$ for $t \in A_j$
1	0–4	0	0
2	5–9	0	0
3	10–14	.1	4.66×10^{-7}
4	15–19	.1	2.38×10^{-8}
5	20–24	1	4.51×10^{-6}
6	25–29	7.5	3.06×10^{-5}
7	30–34	25.2	8.71×10^{-5}
8	35–39	63.8	20.59×10^{-5}
9	40–44	125.4	36.38×10^{-5}
10	45–49	197.8	51.40×10^{-5}
11	50–54	232.7	47.98×10^{-5}
12	55–59	278.0	66.70×10^{-5}
13	60–64	343.3	78.68×10^{-5}
14	65–69	412.1	94.08×10^{-5}
15	70–74	451.0	92.82×10^{-5}
16	75–79	483.9	103.79×10^{-5}
17	80–84	477.4	89.11×10^{-5}
18	85+	432.5	78.78×10^{-5}

NOTES: A_j given in years.

* Rates are from the SEER Program (NCI 1994) per 100,000.

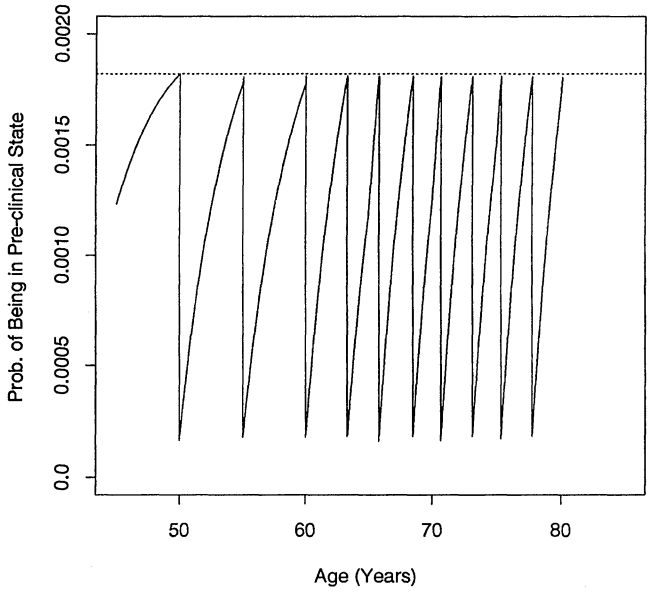

Figure 1. Examination Schedule With $m = 4$, $\beta = .9$, and Age at Initial Examination $t_0 = 50$. Examinations are scheduled whenever the probability of being in the preclinical state reaches the same value as at age 50.

that the exponential distribution was the best fit among a variety of distributions.

We estimated $w(t)$ for a range of mean sojourn times in S_p: $m = 2, 3, 4, 5$. For the given age-specific incidence rate, a larger m leads to a larger $w(t)$. Both age-specific incidence rates and $w(t)$ are step functions with 5-year intervals. Table 1 summarizes the estimated $w(t)$ for a given I_j when $m = 4$.

3.2 Threshold Method

Because the current recommendation for an initial mammogram for women is at age 50, we set the threshold value as the probability of being in S_p at age 50. In computing $P_0(50)$ and $S(t)$, we considered $\beta = .9$ and $m = 4$ as a standard case. This was based on the point estimates supplied by Day, Walter, and Tabar (1988) from the Swedish two-county trial (Tabar, Fagerberg, Gad, Baldetorp, Holmberg, Gromtoft, Ljungquist, Lundstrom, Madson, and Ekhund 1985). For the standard case, we estimated the threshold value $P_0(50) = .0018$. After the first screening at age 50, subsequent examination schedules at ages t_i were obtained so that $S(t_i) = .0018$.

This generated the 10-examination schedule within the age interval [50, 79] of 50, 55.6, 59.8, 62.7, 65.3, 67.8, 70.1, 72.5, 74.9, and 77 years. Note that intervals between the examinations become smaller as women get older. Because incidence is an increasing function of age, the shorter intervals with age agrees with our intuition. Figure 1 plots the probability of being in the preclinical state for women aged [50, 79] who follow this schedule. This schedule allows women to stay under the threshold value of .0018 for the 29-year period. The screening horizon schedule sensitivity (expected fraction of cases diagnosed by a scheduled exam over the screening horizon) is 68.3% for this schedule.

We computed threshold screening schedules for a range of parameters (β, m): $\beta = .8, .9, 1.0$ and $m = 2, 3, 4, 5$. The calculations are summarized in Table 2. It is important to note that the threshold value $P_0(50)$ changes with a specified m. Larger mean sojourn times in S_p are associated with larger $w(t)$, and this leads to larger values of $P_0(50)$. Thus less frequent screening is expected for larger m. A more sensitive screening test (larger β) also requires less frequent examinations. Also, in all cases the intervals between exams decrease with age. Table 2 summarizes the different schedules by calculating the average interval between exams over the 29-year period for different values of (β, m) as well as the number of examinations. The table also includes the average probability of having a disease $(S(t))$ over this period, along with the values of R for the threshold method. Note that R ranges from 63.2% to 73.2% for the screening horizon [50, 79] and that the screening horizon schedule sensitivities increase as β and m increase.

Table 2. Average Exam Intervals, Average Probabilities of Being in Preclinical State ($S(t)$), and Screening Horizon Schedule Sensitivities (R) During the Screening Horizon [50, 79] for Various (β,m)

			β		
m	$P_0(50)$	Age [50, 79]	.8	.9	1.0
2	.0009	Average interval (yr)	1.45	1.53	1.71
		Average $S(t)$	.0006	.0005	.0005
		Total no. of exams	20	19	17
		R (%)	63.2	66.1	67.4
3	.0014	Average interval (yr)	2.07	2.23	2.42
		Average $S(t)$	.0009	.0008	.0008
		Total no. of exams	14	13	12
		R (%)	65.4	67.8	70.0
4	.0018	Average interval (yr)	2.42	2.90	2.90
		Average $S(t)$	.0011	.0010	.0009
		Total no. of exams	12	10	10
		R (%)	68.3	68.3	72.7
5	.0023	Average interval (yr)	2.90	3.22	3.63
		Average $S(t)$	.0014	.0013	.0012
		Total no. of exams	10	9	8
		R (%)	70.0	72.0	73.2

Table 3. Screening Horizon Schedule Sensitivities When the Exam Schedule for (β = .9, m = 4) Is Followed if True Parameters Are (β, m)

	β		
m	.8	.9	1.0
2	45.7	49.9	54.0
3	56.7	60.8	64.9
4	64.2	68.3	72.4
5	66.9	71.0	74.4

NOTE: The screening horizon is [50, 79]; initial age of examination, t_0 = 50. Values given in percentages.

3.3 Robustness

One issue of interest is the robustness of the screening schedules calculated using incorrect values of (β, m). To study this problem for the threshold method, we calculated the schedule sensitivity (R) for the 10-exam schedule based on $(\beta, m) = (.9, 4)$ when (β, m) are actually other values. Table 3 summarizes these schedule sensitivities for $\beta = .8, .9, 1.0$ and $m = 2, 3, 4, 5$. For example, using $(\beta, m) = (.8, 2)$ and the incorrect 10-exam schedule based on $(\beta, m) = (.9, 4)$, the resulting schedule sensitivity was calculated to be 45.7%.

Another way to compare robustness is to calculate the increased number of interval cases due to using the incorrect schedule. For example, according to Table 2, the examination schedule for $(\beta, m) = (.8, 2)$ results in a schedule sensitivity of $R = 63.2\%$ for the same threshold value of $P_0(50) = .0018$. Hence there will be 28% more interval cases $((63.2 - 45.7)/63.2 = .28)$ using the incorrect 10-exam schedule as compared to the schedule based on $(\beta, m) = (.8, 2)$.

Robustness depends on whether the estimated parameters (β, m) induce fewer or larger number of exams compared to the true values. If more exams than necessary are used, then the robustness is good. However, if significantly fewer exams are used, then the change in R may be large, resulting in lack of robustness. Hence it is worthwhile to underestimate both β and m, which builds robustness into the program.

3.4 Comparison of Screening Schedules

Current recommendations for breast cancer screening using mammography (for women aged 50 years and older) vary from every year (United States), every 18–24 months (Scandinavia), and every 3 years (United Kingdom). We compare the characteristics of these schedules with the schedule generated by the threshold method. (It is assumed throughout that $\beta = .9$ and $m = 4$.) Table 4 summarizes the main features of these different strategies in which the screening horizon is [50, 79]. The total number of examinations between ages 50 and 79 varies from 10 to 30. Note that for any age, the probability of being in $S_p(S(t))$ remains under the threshold value of $P_0(50) = .0018$ for the annual and 2-year screening programs, as well as for the threshold method.

The threshold method and screening every 3 years both require 10 exams and for all practical purposes have approximately the same screening horizon schedule sensitivity. However, the 3-year screening schedule does not result in the probability $S(t)$ being less than the threshold value of .0018 after age 60. The other strategies of screening every 1 or 2 years give higher schedule sensitivities because they involve more examinations. Note that the annual exam schedule will diagnose $R = 87.2\%$ of the cases with 30 exams, whereas the threshold method will result in a value of $R = 68.3\%$ with 10 exams. Hence the threshold method can detect 78% $(63.8/87.2 = .78)$ of the cases that could have been detected by annual screening but requires only one-third of the total number of exams. Table 4 also compares the cost per case diagnosed (8) for the 1- and 2-year strategies relative to the threshold method. Note that the costs per case diagnosed are substantially higher for the 1- and 2-year screening schedules.

3.5 Examination Schedules for High-Risk Individuals

Often individuals are at higher risk if they have a familial association with a particular chronic disease. For breast cancer, women may be at higher risk due not only to their family history, but also to other factors such as age at menarche, parity, and age at first live birth (Gail, Brintom, Byar, Corle, Green, Schairer, and Mulvihill 1989). Recent research has identified possible links between heritable defects in specific genes and increased risk of such cancers as nonpolyposis colon cancer (Fishel, Lescoe, Rao, Copeland, Jenkins, Garber, Kane, and Kolodner 1993) and breast and ovarian cancer (Struewing, Hartge, Wacholder, Baker, Berlim, McAdams, Timmerman, Brody, and Tucker 1997). Women carrying the BRCA1 gene, the first gene discovered for in-

Table 4. Comparison of Different Schedules (β = .9, m = 4)

Total no. of exams	Threshold 10	Annual 30	Every 2 yr 15	Every 3 yr 10
Screening horizon schedule sensitivity (R) (%)	68.3	87.2	77.1	68.9
Relative cost per case found on scheduled exam	1.00	2.24	1.30	0.97
Age interval (yr)	*Range of S(t) at the Time of Screening*			
50–60	≤.0018	.0005–.0006	.0008–.0011	.0011–.0019
60–70	≤.0018	.0007–.0009	.0013–.0016	.0017–.0024
70–79	≤.0018	.0009–.0010	.0016–.0018	.0021–.0026

NOTE: The screening horizon is [50, 79] and initial age of examination is t_0 = 50.

Table 5. Screening High-Risk Groups (β = .9, m = 4) Keeping the Cost per Case to be the Same Independent of Risk Status

	Risk				
	1	1.5	2	3	4
Initial age (t_0)	50	45	41	38	36
Age interval (yr)	No. of exams				
30–40	0	0	0	1	2
40–50	0	2	4	7	10
50–60	3	4	6	9	13
60–70	3	7	9	16	22
70–79	4	7	10	16	23
Total no. of exams	10	20	29	49	70
Average $S(t)$	.0011	.0010	.0010	.0010	.0009
Lifetime schedule sensitivity (R_0) for age [0, 79] (%)	58.0	73.2	80.0	87.0	90.5
Relative cost per case found on scheduled exam ($)	1.00	1.00	1.00	1.00	1.00

herited breast cancer, have an estimated 40%–73% lifetime risk for breast cancer and a 6%–28% risk for ovarian cancer. Such genetic discoveries are likely to continue, leading to identification of more high-risk populations.

Individuals at high risk need special attention. Although there are no established guidelines for screening schedules for high-risk individuals, it seems necessary to have a schedule with an earlier initial examination followed by more frequent subsequent examinations relative to the normal-risk population. In Section 2.4 we proposed a method for calculating screening programs under the assumption of elevated age-specific incidence rates. For example, if the risk is elevated by a factor of two, then the age-specific incidence rates are multiplied by two across all age groups in the computation of $w(t)$. The resulting $w(t)$ and $S(t)$ are larger at any given age t compared to the population not at higher risk. Any appropriate assumption on β or m can be used. For the purposes of illustration and comparison, we use $\beta = .9$ and $m = 4$ for our calculations.

One possible way to derive the screening schedule for a high-risk group is to find the age at which $S(t)$ first reaches the threshold value of .0018. This is the value of $S(t)$ for the normal-risk population at age $t_0 = 50$ and determines the age of the initial exam. Subsequent examination times are found to keep $S(t)$ under .0018. Table 5 summarizes the characteristics of screening programs for risk ratios ranging from 1 to 4. With the risk ratio of 2, for example, the recommended screening schedule requires 29 exams at ages 41, 44.6, 46.7, 48.8, 50.6, 52.9, 55.1, 56.6, 58.1, 59.6, 60.8, 61.9, 63.1, 64.3, 65.3, 66.3, 67.2, 68.2, 69.1, 70, 71, 72, 73, 74, 75, 75.8, 76.7, 77.5, and 78.3. This is in contrast to the comparable program for the normal-risk population having 10 exams, which is given in Section 3.2. Note that even when the risk is elevated by a factor of two, the spacing between exams exceeds 1 year until age 64.3 years. Average values of $S(t)$ during the recommended follow-up period remain between .0009 and .0011 for the high-risk groups. This value of $S(t)$ is approximately the same as the average value of .0011 for the non–high-risk population. To adjust for the different initial ages, the lifetime schedule sensitivities (R_0) were computed for these high-risk individuals. More frequent examination schedules provide larger values of R_0 (73.2% to 90.5%) in this population compared to the normal-risk population (58.0%). Note that the schedule was chosen so that the cost per case found is the same regardless of risk status.

Alternatively, the screening schedule for the high-risk group can be generated based on a prespecified R_0. Because R_0 is 58% for the normal-risk population, we can limit the possible high-risk screening schedules to those for which $R_0 = 58\%$. In this strategy, the initial age of screening is chosen such that the cumulative incidence probability I_0 in (5) is approximately the same as that of the normal-risk individuals who start screening at age 50. According to our calculations, the normal-risk population has a probability of being incident of .0042 within the age interval (0, 50). To have the same probability of being incident before the initiation of screening, the high-risk population (with elevated risk of 1.5–4) needs to start screening at age 40–45. Using these initial ages, we chose threshold values iteratively to achieve the prespecified $R_0 = 58\%$. Table 6 summarizes the characteristics of the screening schedules generated by this approach. For the elevated risk of 1.5–4, the threshold values exceed .0018 and range from .0032 to .0087. Furthermore, by initiating screening at earlier ages, 10-examination schedules generated by the threshold method gives a R_0 of about 58% regardless of the magnitude of the risk. Also, the cost per case diagnosed at a scheduled examination is substantially reduced relative to the normal-risk population. Thus the high-risk population can achieve the same R_0 provided that the characteristics (sensitivity and mean sojourn time) are the same as for the normal-risk population.

4. PLANNING SCREENING PROGRAMS WITH PREDETERMINED SCHEDULE SENSITIVITIES

As described earlier (Sec. 2.5), one way of characterizing a screening program is to evaluate the schedule sensitivity over the screening horizon or over the lifetime. It is clearly a function of the initial age at screening, probability of being in S_p at an exam, sojourn time distribution, total number

Table 6. Screening High-Risk Groups ($\beta = .9$, $m = 4$) With Ten Exams

	Risk				
	1	*1.5*	*2*	*3*	*4*
Initial age (t_0)	*50*	*44*	*43*	*42*	*40*
Threshold value	*.0018*	*.0032*	*.0040*	*.0065*	*.0087*
Age interval (yr)			*No. of exams*		
40–50	0	1	1	2	2
50–60	3	2	2	1	1
60–70	3	3	3	3	3
70–79	4	4	4	4	4
Total no. of exams	10	10	10	10	10
Average $S(t)$	.0011	.0018	.0025	.0037	.0050
Lifetime schedule sensitivity (R_0) for age [0, 79] (%)	58.0	60.4	57.2	58.2	59.2
Relative cost per case found on scheduled exams	1.00	.59	.58	.30	.22

NOTE: Lifetime schedule sensitivities are not exactly 58%, as the initial ages (t_0) are confined to whole numbers.

of examinations, and sensitivity of the screening test. In designing a screening program, this relationship can be used to analytically determine the initial age to begin screening, as well as the frequency of the subsequent examinations for specified β, m, and R or R_0.

Schedules of screening exams can be generated by specifying in advance the screening horizon schedule sensitivity. But prespecifying the screening horizon schedule sensitivity does not necessarily result in unique schedules. For example, if the screening horizon schedule sensitivity is fixed, then one can lower or raise the value of the threshold probability after the initial examination and find schedules with the same schedule sensitivity. To illustrate this phenomenon for breast cancer, suppose that $R = 80\%$ and the screening horizon is [50, 79]. Then after the first exam at age 50, one can choose the threshold probability to be .0012. This schedule will result in $R = 80\%$ and requires 18 examinations. (These calculations assume that $\beta = .9$ and $m = 4$.) Alternatively, a schedule sensitivity of $R = 80\%$ can be attained by lowering the initial age to $t_0 = 44$ and using $P_0(t_0) = .0011$ as a threshold probability. This strategy requires 22 examinations for the screening horizon [44, 79]. The relative cost per case found comparing the 22-exam schedule to the 18-exam schedule is 1.18, showing that the 18-exam schedule is superior in cost per case found. Screening schedules based on the prespecified lifetime schedule sensitivities account for the incident cases before the age t_0. Table 7 summarizes our calculations of examination schedules with $R_0 = 75\%$ and $t_0 = 40, 45, 50$. It is clear that a lower threshold value and a more frequent examination schedule are needed to achieve a higher R_0. It is interesting to note that by initiating screening at earlier ages, one can achieve an $R_0 = 75\%$ with a smaller number of total examinations within the interval [0, 79].

Table 7. Screening Schedules for Predetermined Lifetime Schedule Sensitivities (R_0) of 75% for Age Interval [0, 79]

Initial age (yr)	*Threshold value*	*Total no. of exams*	*R_0 with $\beta = .90$ (%)*
50	.0008	30	75.3
45	.0011	22	74.6
40	.0012	21	74.6

5. DISCUSSION AND SUMMARY

5.1 General Remarks

Diagnosing disease at an early stage offers the best opportunity for cure or improved prognosis with many chronic diseases. This is especially true for nearly all solid-tumor cancers, where the prognosis for cure is higher if the disease is localized at the time of detection and can be excised by surgery. Actually there are no cancer sites in which populations of individuals with localized disease do not fare better with primary treatment compared to those diagnosed with non-localized disease. The issue for benefit is not finding disease earlier, but finding a higher proportion of curable disease cases by earlier detection.

Calculation of schedule sensitivities requires data on age-specific incidence of disease, sensitivity of the screening modalities, and distribution of sojourn times in the preclinical state. Age-specific incidence data are available for most chronic diseases and are key to all calculations. Sensitivity (or a range of sensitivities) is available for any proposed special examination procedures. Less readily available is the sojourn time distribution in the preclinical state. It is possible (as we have illustrated) to adopt a range of sojourn time distributions to investigate the consequence for the screening program. In our illustrations for breast cancer screening, we have used the exponential distribution, but the theoretical results are general and any probability distribution can be used.

Some would argue that the screening programs should be assessed by survival benefit. It should be noted that for detection modalities of proven benefit, any ordering of screening programs by survival benefit will be the same as the

ordering according to the screening horizon schedule sensitivities. However, the principal reason for not using survival data in the model is that such data can come only from early detection trials. Yet these data are generally unreliable and biased due to significant noncompliance problems. Furthermore, patient treatment after diagnosis in an early detection program is unknown and may vary between hospitals as well as between trials.

5.2 Calculation and Evaluation of Schedules

We have introduced two new ideas for the calculation of early examination schedules: the threshold method and prespecified schedule sensitivities. The threshold method guarantees that the probability of an individual being in the preclinical state is always bounded. This method is particularly appropriate for breast cancer screening, as there is a universal recommendation that women should begin mammogram examinations at age 50. Because breast cancer incidence increases with age, if a woman did undergo a single exam at age 50 and no other, then eventually the probability of being in the preclinical state would exceed that value at age 50. However, utilizing the threshold method allows a woman to have the same or lower risk status at more advanced ages even though the risk increases with age.

Another way of determining an examination schedule is to set the screening horizon schedule sensitivity at a predetermined value. This method does not result in unique schedules. An open problem is to determine the schedules using a minimum number of exams for a fixed screening horizon schedule sensitivity.

The schedule sensitivity and relative costs can be used to evaluate the consequences of any screening schedule. Choosing an optimal screening program requires balancing outcomes (schedule sensitivity) and costs. The two principal measures of a screening program in our development are the schedule sensitivities (screening horizon or lifetime) and cost. Any selection of screening schedules can be ordered with respect to the proportions of individuals diagnosed on scheduled examinations. The selection of one of these schedules for implementation in a public health program is a policy issue, as it will depend on trade-offs between benefit and costs.

5.3 Breast Cancer Screening

We have applied our theoretical models to calculate breast cancer screening examination times for women aged 50 and older. We have arbitrarily taken the screening horizon to be [50, 79]. There are approximately 42,000,000 women in the United States aged 50 and older. If a special breast cancer examination, which includes a mammogram and a physical examination, costs \$100–200, and if all women undergo such an examination, the total annual cost would be \$4.2–8.4 billion. As noted earlier, recommendations for breast cancer screening vary among nations. Presumably, all recommendations have been based on the same scientific data. Disagreements about schedules is reflected in billions of dollars in health care costs.

Most United States women do not receive annual mammogram examinations. The difference between the 10-exam threshold schedule and the 15-exams-every-2-year schedule results in a 11% increase in interval cases $((77.1 - -68.3)/77.1 = .11)$ with a 50% reduction in the cost of screening; that is, 10 exams versus 15 exams. The decision as to which schedule to use is clearly a health policy decision, as the 11% increase in interval cases results in a savings of \$20–40 billion over a 30-year period. (We mention these dollar figures simply to show that billions of dollars are available for cost savings.)

We have purposely avoided the issue of screening women under age 50 for breast cancer. The problem is complicated and requires a detailed discussion of each of the clinical trials. We plan to submit an article on this issue sometime in the future.

5.4 Extensions

Many extensions to our basic model can be made in a straightforward way. We mention three extensions that can be easily done and may be important in applications.

In some instances, sensitivities and mean sojourn time may be functions of age. In this case we can take both quantities $\beta(t)$ and $m(t)$ to be functions of age. The modifications are easily incorporated into our general model. A future modification is that the mean sojourn time may also depend on whether the risk status is elevated. Another extension is that two or more detection modalities with different sensitivities may be used. These can also be incorporated into the model.

Finally, we conclude our discussion with some remarks on models. No model is really correct, because the observed phenomena usually are much more complex than can be incorporated into a model. Nevertheless, models can incorporate the main features of the observed phenomena and can lead to guides for action. This is the spirit in which we have put forth our models to plan public health programs for the early diagnosis of chronic disease.

[Received November 1996. Revised February 1998.]

REFERENCES

Albert, A., Gertman, P. M., Louis, T., and Liu, S. (1978), "Screening for the Early Detection of Cancer II. The Impact of Screening in the Natural History of the Disease," *Mathematical Bioscience*, 40, 61–109.

Baker, S. G., and Chu, K. C. (1990), "Evaluating Screening for the Early Detection and Treatment of Cancer Without Using a Randomized Control Group," *Journal of the American Statistical Association*, 85, 321–327.

Day, N. E., and Walter, S. D. (1984), "Simplified Models of Screening of Chronic Disease: Estimation Procedures From Mass Screening Programmes," *Biometrics*, 40, 1–13.

Day, N. E., Walter, S. D., and Tabar, L. (1988), "The Sensitivity and Lead Time of Breast Cancer Screening: A Comparison of the Results of Different Studies," in *Screening for Breast Cancer*, eds. N. E. Day and A. B. Miller, Stuttgart: Hans Huber, pp. 105–109.

Dubin, N. (1981), "Predicting the Benefit of Screening for Disease," *Journal of Applied Probability*, 18, 348–360.

Eddy, D. M. (1980), *Screening for Cancer: Theory, Analysis and Design*, Englewood Cliffs, NJ: Prentice-Hall.

——— (1983), "A Mathematical Model for Timing Repeated Medical Tests," *Medical Decision Making*, 3, 34–62.

Eddy, D. M., and Shwartz, M. (1982), "Mathematical Models in Screening," in *Cancer Epidemiology and Prevention*, eds. W. Schottenfeld and J. Fraumeni, Philadelphia, PA: W. B. Saunders, pp. 1075–1090.

Fishel, R., Lescoe, M. K., Rao, M. R., Copeland, N. G., Jenkins, N. A., Garber, J., Kane, M., and Kolodmer, R. (1993), "The Human Mutation Gene Homolog MSH2 and Its Association With Nonpolyposis Colon Cancer," *Cell*, 75, 1027–1038.

Fletcher, S. W., Black, W., Harris, R., Rimer, B. K., and Shapiro, S. (1993), "Report of the International Workshop on Screening for Breast Cancer," *Journal of the National Cancer Institute*, 85, 1644–1656.

Gail, M. H., Brinton, L. A., Byar, D. P., Corle, D. K., Green, S. B., Schairer, C., and Mulvihill, J. J. (1989), "Projecting Individualized Probabilities of Developing Breast Cancer for White Females Who Are Being Examined Annually," *Journal of the National Cancer Institute*, 81, 1879–1886.

Habbema, J. D. F., VanOortmarssen, G. J., Lubbe, J. T. N., and van der Mass, P. J. (1984), "The MISCAN Simulation Program for the Evaluation of Screening for Disease," *Computer Programs in Biomedicine* (Vol. 9), Amsterdam, North-Holland.

Kirch, R. L. A., and Klein, M. (1974), "Examination Schedules for Breast Cancer," *Cancer*, 33, 1444–1450.

Knox, E. G. (1973), "A Simulation System for Screening Procedures," in *The Future and Present Indicatives, Problems and Progress in Medical Care, Ninth Series, Nuffield Provincial Hospitals Trust*, ed. G. McLachlan, London: Oxford University Press, pp. 17–55.

National Cancer Institute (1994), *Surveillance, Epidemiology, and End Results Program. Age-Specific Rates for Breast Cancer*. Bethesda, MD: Author.

NIH Consensus Statement (1997), *Breast Cancer Screening for Women Ages 40–49*, Bethesda, MD: National Cancer Institute.

Parmigiani, G. (1993), "On Optimal Screening Ages," *Journal of the American Statistical Association*, 88, 622–628.

——— (1997), "Timing Medical Examinations via Intensity Functions," *Biometrika*, 84, 803–816.

Prorok, P. C. (1976a), "The Theory of Periodic Screening I. Lead Time and Proportion Detected," *Advanced Applied Probability*, 8, 127–143.

——— (1976b), "The Theory of Periodic Screening II. Doubly Bounded Recurrence Times and Mean Lead Time and Detection Probability Estimation," *Advanced Applied Probability*, 8, 460–476.

Shwartz, M. (1978), "An Analysis of the Benefits of Serial Screening for Breast Cancer Based Upon a Mathematical Model of the Disease," *Cancer*, 41, 1550–1564.

Shwartz, M., and Plough, A. L. (1984), "Models in Planning Cancer Programs," in *Statistical Methods for Cancer Studies*, ed. R. G. Cornell, New York: Marcel Dekker, pp. 329–416.

Shapiro, S., Venet, W., Strax, P. H., Venet, L., and Rosener, R. (1982), "Ten- to Fourteen-Year Effect of Screening on Breast Cancer Mortality," *Journal of the National Cancer Institute*, 69, 349–355.

Struewing, J. P., Hartge, P., Wacholder, S., Baker, S. M., Berlim, M., McAdams, M., Timmerman, M. M., Brody, L. C., and Tucker, M. A. (1997), "The Risk of Cancer Associated With Specific Mutations of BRCA1 and BRCA2 Among Ashkenazi Jews," *New England Journal of Medicine*, 11, 1401–1408.

Tabar, L., Fagerberg, C. J., Gad, A., Baldetorp, L., Holmberg, L. H., Grontoft, O., Ljungquist, U., Lundstrom, B., Mansom, J. C., and Ekhund, G. (1985), "Reduction in Mortality From Breast Cancer After Mass Screening With Mammography," *Lancet*, 1, 829–832.

Zelen, M. (1993), "Optimal Scheduling of Examinations for the Early Detection of Disease," *Biometrika*, 80, 279–293.

Zelen, M., and Feinleib, M. (1969), "On the Theory of Screening for Chronic Diseases," *Biometrika*, 56, 601–614.