AD_____

Award Number: DAMD17-98-1-8256

TITLE: Individualized Strategies for Breast Cancer Surveillance Based on Aggregated Familial Information

PRINCIPAL INVESTIGATOR: Andrei Y. Yakovlev, Ph.D.

CONTRACTING ORGANIZATION: University of Utah
Salt Lake City, Utah 84102

REPORT DATE: January 2000

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	January 2000	Annual (1 Jan 99 - 31 Dec 99)

4. TITLE AND SUBTITLE
Individualized Strategies for Breast Cancer Surveillance Based on Aggregated Familial Information

5. FUNDING NUMBERS
DAMD17-98-1-8256

6. AUTHOR(S)
Andrei Y. Yakovlev, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Utah
Salt Lake City, Utah 84102
E-MAIL: yak@hci.utah.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 Words)*

This work is concerned with methodological aspects of a new approach to optimization of breast cancer screening designed to utilize aggregated family history information. The problem of optimal cancer surveillance is set up as a search for optimal scheduling of screening examinations subject to certain constraints on the number and timing of medical tests. In Year 1, we developed a mathematical model yielding an algorithm for designing optimal schedules of breast cancer screening. An explicit expression of the efficiency functional is based on a plausible assumption that the intensity of detection (the hazard function for the age at detection) is proportional to the current tumor size. The main advantage of the proposed approach is that it accommodates cohort data of a fairly general structure, not only the data resulting from screening trials. We have also developed several numerical algorithms and software for estimating the hazard function for breast cancer incidence from the data amassed in the Utah Population Data Base; these procedures will be used (in Year 2) for testing covariate effects associated with different indicators of family history.

14. SUBJECT TERMS
Breast Cancer

15. NUMBER OF PAGES
39

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

N/A Where copyrighted material is quoted, permission has been obtained to use such material.

N/A Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

N/A Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

N/A In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

X For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

N/A In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

N/A In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

N/A In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

01/12/00

PI - Signature Date

Table of Contents

Introduction

This work is concerned with theoretical methods for designing individualized optimal strategies of breast cancer surveillance. The problem of optimal cancer surveillance is set up as a search for optimal scheduling of screening examinations subject to certain constraints on the number and timing of medical tests. The hypothesis to be tested is that the efficacy of breast cancer detection can be enhanced through incorporating aggregated family history information into a mathematical model designed to construct optimal schedules of cancer surveillance. The proposed methods are to be validated using epidemiological data on breast cancer from the Utah Population Data Base and the Utah Cancer Registry.

1. Statement of Work

This annual report covers the following four tasks formulated in the statement of work.
Task 1: Develop programs for estimating the hazard function for breast cancer using the data-adaptive method of kernel estimation.
Task 2: Develop programs for estimating the hazard function for breast cancer using the spline approximation method.
Task 3: Develop programs for estimating the hazard function for breast cancer using parametric estimation techniques.
Task 4: Construct optimal surveillance scheduling strategies based on the three methods of hazard function estimation.

2. The research carried out to meet the objectives of Tasks 1 and 2

2.1. Introduction

The shape of the hazard function may lead to insights into the biology of carcinogenesis which may not be easily discernable from a study of the survival function alone. For example, it is typical in the analysis of tumor recurence data to find a hazard function that is bimodal or unimodal, and that tends to zero as time tends to infinity [1]. The modes of the hazard may be interpreted biologically as arising from two different types of failure, one that tends to occur earlier and one that tends to occur later. In the context of the age-specific hazard function for cancer incidence the decrease in the hazard function to zero may lead one to conclude that there is a non-zero fraction of unaffected (immune) individuals. In fact, if the cumulative hazard appears to be bounded, one should expect the existence of a non-zero immune fraction. More generally, a large degree of heterogeneity in disease susceptibility may lead to a population hazard function with one or more well-defined maxima. The maxima may correspond to discrete subpopulations with different genetic predisposition to disease. A maximum may also result from a continuous frailty, as

the surviving population at higher ages may be overrepresented by individuals with lower risk [2]. On the other hand, a monotonically increasing (with age at detection) hazard rate is consistent with a popular belief that spontaneous tumorigenesis can be modeled as a Poisson process.

The form of the hazard function for breast cancer incidence may depend on characteristics inherited susceptibility. Inherited mutations in p53, BRCA1, BRCA2, the ataxia-telangiectasia gene (AT), HRAS, and the androgen receptor gene (AR) have been shown to play a role in breast cancer susceptibility [3]. About 56% of carriers of the mutation BRCA1 or BRCA2 will get breast cancer by the age of 70 years [4]. BRCA1 has an estimated allele frequency of between 0.0002 and 0.001 (95% CI) [5], and accounts for about 3% of diagnosed breast cancer [6]. The allele frequency of mutations in BRCA2 is estimated at 0.00022 [7]. Germline mutations in p53 and AR are extremely rare, and mutations in the HRAS1 minisatellite locus which confer increased risk of breast cancer are also rare, having an estimated population frequency of 6% [3]. In a study of 100 Finnish breast cancer families analyzed by protein truncation tests and direct sequencing, Vehmanen et al. [8] found that only 21% of breast cancer families were accounted for by mutations of BRCA1 and BRCA2, providing indirect evidence for the existence of other, undiscovered breast cancer genes.

Additional insight can be gleaned from the hazard function for cancer incidence in the framework of a mechanistic model of carcinogenesis. The most widely accepted model is the Moolgavkar-Venzon-Knudson two-stage clonal expansion model [9,10]. The Moolgavkar-Venzon-Knudson model has the following assumptions:
(A) Normal, susceptible target cells are initiated according to a (nonhomogeneous) Poisson process with intensity $\nu(t)$.
(B) The expansion of the colony of initiated cells and malignant transformation is specified by a stochastic birth-death-migration process with the division, death (or differentiation) and transformation. Premalignant cells either divide into two premalignant cells with rate $\alpha(t)$, die with rate $\beta(t)$, or divide asymmetrically into one premalignant cell and one malignant cell with rate $\mu(t)$.

It has been shown that the hazard function for the Moolgavkar-Venzon-Knudson model with constant parameters increases monotonically and approaches an asymptote [11]. An asymptotic value for the hazard is also reached for the Moolgavkar-Venzon-Knudson model with piecewise constant parameters, and in that case the value of the asymptote depends only on the value of the coefficients in the unbounded interval [11,12].

Expressions for the survivor function were first obtained by Moolgavkar and Luebeck [11]. A simple explicit formula for the survivor function $S(t)$ for the Moolgavkar-Venzon-Knudson model with constant parameters was obtained by Kopp-Schneider et al. [13] and Zheng [14]:

$$S(t) = \left[\frac{2ce^{0.5(-\alpha+\beta+\mu-c)t}}{(-\alpha+\beta+\mu+c)+(\alpha-\beta-\mu+c)e^{-ct}} \right]^{\nu/\alpha} \tag{1}$$

where $c = \sqrt{(\alpha+\beta+\mu)^2 - 4\alpha\beta}$. Zheng also presented an expression for the proba-

bility generating function for the number of malignant cells given a single malignant cell at time $t = 0$, allowing an expression for the promotion time distribution

$$F(t) = \frac{(\alpha - \beta - \mu + c)(\alpha - \beta - \mu - c)e^{-ct} + (\alpha - \beta - \mu + c)(-\alpha + \beta + \mu + c)}{2\alpha[(\alpha - \beta - \mu + c)e^{-ct} + (-\alpha + \beta + \mu + c)]} \tag{2}$$

to be given. It is easy to see that $S(t)$ and $F(t)$ above are related by the formula

$$S(t) = \exp\left\{-\nu \int_0^t F(x)dx\right\} \tag{3}$$

which was shown by Hanin and Yakovlev [15] to be valid in a more general setting.

Yakovlev and Tsodikov [16] replace assumption (B) above with the following assumption:
(C) Progenitor cells are transformed into malignant lesions at a random with cumulative distribution function $F(x)$. All progenitor cells are promoted independently of one another.
Assuming $F(0) = 0$, it follows that the process of malignant transformation is also a Poisson process, with integral rate $\Lambda(t) = \int_0^t \nu(u)F(t-u)du$. As in the Moolgavkar-Venzon-Knudson model, the simplest model of spontaneous carcinogenis takes $\nu(t) = \nu$ to be constant, in which case $\Lambda(t) = \nu \int_0^t F(u)du$ and the hazard function for time-to-tumor, given by $\lambda(t) = \nu F(t)$, is nondecreasing. The probability $S(t)$ that there are no malignancies by time t is then given by (3).

This model may easily be modified to handle inherited lesions, via the limiting case where ν is taken to be a delta function at the origin. If $F(t)$ is assumed to be absolutely continuous, then the integral hazard rate $\Lambda(t)$ is equal to $\nu F(t)$ and the hazard function $\lambda(t) = \nu F'(t) = \nu f(t)$, where $f(t)$ is the probability density function associated with $F(t)$. We see that the hazard function for spontaneous and inherited lesions are quite likely to have very different shapes.

Even though a thorough study of the hazard function may lead to new insight into the process of carcinogenesis, few if any population-based cohorts have been analyzed to determine the hazard function for cancer incidence.

2.2. Data

The data for this study was obtained by linking records from the Utah Population Data Base (UPDB) with the Utah Cancer Registry (UCR). The UPDB consists of the genealogical records of more than 1,000,000 individuals who were born, died, or married in Utah, or en route to Utah during the nineteenth and twentieth centuries. Since 1973 the UCR has been reporting to National Cancer Institutes Surveillance Epidemiology and End Results (SEER) program, and is required to maintain very high standards for case reporting and follow-up, and to periodically undergo quality control audits by SEER personnel to assure uniformly high quality and consistency from year to year. The available follow-up information comes either from Utah death certificates, which have been linked to the UPDB genealogical data every year from 1933 through the beginning of 1997, or from linkage of the HCFA beneficiary data to the UPDB. The study population consists of 126,141 men and 122,208 women

recorded in the Utah Population Database, who were born from 1874 to 1931 and for whom follow-up information is available that places them in Utah during the years of operation of the Utah Cancer Registry (1966-present). There are 5,372 cases of female breast cancer represented in the data. Since methodological research was the major undertaking in Year 1, we plan to conduct analyses of epidemiological data in 2000. Analyses will be performed on subcohorts based on birth year (1874-1889, 1890-1899, 1900-1909, 1910-1919, and 1920-1931).

2.3. Methodological Problems

Truncation: Nonparametric Estimation

We wish to estimate the age specific hazard function for breast cancer from the data described above, taking into account that the data is subject to random truncation: cases which occurred during or before 1965 are not recorded in the dataset. Subject were between the ages of 34 and 86, at the time of truncation. Thus, analysis of the data must take into account not only to the effects of right censoring, but also the effects of left random truncation due to delayed entry into the risk set.

The problem of random truncation can be formulated as follows. Let the truncation time Y have distribution function $G(y)$ and the failure time (time of cancer diagnosis) X have distribution function $F(x)$. We require that truncation be independent of failure and for simplicity assume no censoring for the present. Observations are conditional on the random event $X > Y$. Let $G^*(y)$ and $F^*(x)$ be the corresponding distribution functions, conditional on $X > Y$. Let $S(x) = 1 - F(x)$ be the survivor function of X. Suppose that we have observations $(Y_1^*, X_1^*), \ldots, (Y_n^*, X_n^*)$ from the conditional distribution. The full likelihood of the observed data is given by

$$L = \prod_{j=1}^{n} \left[dF(X_j) dG(Y_j)/\alpha \right], \tag{4}$$

where $\alpha = \int \int_{y \leq x} dF(x) dG(y)$. A key observation is that if X and Y are independent, then the hazard of X given $X > Y = y$ at $x > y$ is equal to the hazard of X at x [17,18]. This observations leads to the result, first mentioned by Kaplan and Meier [19], that if the distribution $G(t)$ is allowed to vary freely, the natural generalization of the product limit estimator, given by the formula

$$\hat{S}(t) = \prod_{X_i^* \leq t} \left(1 - \frac{1}{R(X_i^*)} \right), \tag{5}$$

where $R(u) = \#\{Y_i^* < U \leq X_i^*\}$ is the number of subjects at risk at U, is the nonparametric maximum likelihood estimator (NPMLE) of the survivor function $S(t)$ of X (see, for example [17,18,20]).

This result extends naturally to the case with random independent censoring [20]. It also easily follows that in the nonparametric setting (again with no censoring), maximizing (1) is equivalent to maximizing the conditional likelihood of $(X_1^*, \ldots, X_n^*)$

given $(Y_1^*, \ldots, Y_n^*)$, which can be written

$$CL = \prod_{i=1}^{n} f(X_i^*)/S(Y_i^*). \tag{6}$$

(see, for example, [19-22]). Maximizing the conditional likelihood also leads to the familiar Nelson-Aalen estimator for the integrated hazard function $H(t)$ of X [20], which is given by

$$\hat{\Lambda}(t) = \sum_{X_i^* \leq t} R(X_i^*)^{-1}. \tag{7}$$

These results can be extended to the case of right censoring [20].

Truncation: Parametric Models

We consider the situation where X and Y are independent, $F(x)$ is parametrized, while $G(y)$ is allowed to vary freely. In a later subsection $F(x)$ will be come from a quadratic spline model.

The data are independent pairs $(y_1, x_1), \ldots, (y_n, x_n)$ from the joint distribuition (Y, X), conditional on $(Y < X)$. We suppose, for simplicity, that there are no ties among $y_1, y_2, \ldots, y_n$, and suppose X has absolutely continuous distribution function coming from a family $F(x; \vec{z})$ parameterized by a vector $\vec{z}$, with corresponding survival function $S(x; \vec{z}) = 1 - F(x; \vec{z})$ and density $f(x; \vec{z})$. The NPMLE for G should consist of (unknown) point masses $q_1, q_2, \ldots, q_n$ placed at the points $y_1, y_2, \ldots, y_n$. The logarithm of the complete likelihood (4) can be rewritten

$$\log(L) = \sum_{i=1}^{n} [\log(f(x_i; \vec{z})) + \log(q_i)] - n \log \left[\sum_{j=1}^{n} S(y_i; \vec{z}) q_i \right]. \tag{8}$$

If we factor the out the part of the likelihood corresponding to formula (6), the logarithm is given by

$$\log(CL) = \sum_{i=1}^{n} [\log(f(x_i; \vec{z})) - \log(S(y_i; \vec{z}))]. \tag{9}$$

We now discuss the changes which must be made in when censoring and additional covariates are present. If $\vec{s}$ is a vector of additional covariates, $\lambda(x, \vec{s}; \vec{z})$ denotes the hazard associated with $F(x, \vec{s}; \vec{z})$ and $\Lambda(x, \vec{s}; \vec{z})$ the cumulative hazard, we note that (9) becomes

$$\log(CL) = \sum_{i=1}^{n} [\log(\lambda(x_i, \vec{s}; \vec{z})) - (\Lambda(x_i, \vec{s}_i; \vec{z}) - \Lambda(y_i, \vec{s}; \vec{z}))]. \tag{10}$$

In the presence of right censoring which is independent of both the failure and truncation times, x_i is replaced in the above formulation by the minimum of the failure and censoring time. The term $f(x, \vec{s}; \vec{z})$ in the likelihood is replaced by

$f(x, \vec{s}; \vec{z})^{\delta} S(x, \vec{s}; \vec{z})^{1-\delta}$, where $\delta_i = 1$ if observation i is a failure and $\delta_i = 0$ otherwise, and the conditional likelihood (6) (with $x_i, \vec{s}_i$ and y_i regarded as fixed) becomes

$$CL = \prod_{i=1}^{n} [f(x_i, \vec{s}_i; \vec{z}_i)^{\delta} S(x_i, \vec{s}_i; \vec{z}_i)^{(1-\delta)}] / S(y_i, \vec{s}_i; \vec{z}_i).$$

In this setting $\log(CL)$ becomes

$$\log(CL) = \sum_{i=1}^{n} [\delta_i \log(\lambda(x_i, \vec{s}_i; \vec{z}_i)) - (\Lambda(x_i, \vec{s}_i; \vec{z}) - \Lambda(y_i, \vec{s}_i; \vec{z}_i))]. \tag{11}$$

W havee chosen to maximize (11) rather than the full likelihood. Under appropriate conditions the two maximization procedures are equivalent.

Spline Models

We choose to model the hazard via quadratic splines as in [23]. A quadratic spline with m knots specifies the hazard to be of the form

$$\lambda_m(t) = \sum_{i=0}^{2} \gamma_{0i} t^k + \sum_{j=1}^{m} \gamma_{j2} (t - \tau_j)_+^2 \tag{12}$$

where $(x)_+ = \max(x, 0)$. For each birth cohort, we will fit splines with knots which are equally spaced in the interior of the interior $[T_{min}, T_{max}]$, where T_{min} is the minimum truncation age in the cohort and T_{max} the maximum follow-up (failure or censoring) time. Restrictions should be placed on the coefficients to ensure that $\lambda_m(t)$ remains positive for all t. Thus with m knots the number of parameters is $m + 3$. Models can be fit using maximum likelihood techniques applied to the conditional likelihood, as given by (11).

We have developed software designed to compute the spline estimates by maximizing $\log(CL)$ using the algorithm of Powell [24]. We start with one knot and increase the number of knots until the fit is not improved, as determined by the likelihood ratio test at the significance level $\alpha = 0.05$. Two other subcohort estimates of the hazard function can be computed for comparison with the spline estimator; a life table version of (5), and a Gaussian kernel estimate based on the Nelson-Aalen estimator (7). We have developed relevant software for this purpose as well.

2.4. Future Plans

Using the computer programs developed in Year 1, the hazard function for cancer incidence will be estimated from left truncated and right censored data based on the the conditional likelihood. Four estimation procedures based on the conditional likelihood will be used to estimate the age-specific hazard function from the data; these are the life-table method, a kernel method based on the Nelson Aalen estimator, a spline estimate, and a proportional hazards estimate based on splines with birth year as sole covariate. The latter model is aimed at adjusting for the cohort effect.

3. The research carried out to meet the objectives of Tasks 3 and 4

3.1. Introduction

When evaluating potential benefits of screening in different populations (different racial composition, different geographic regions, etc.), it makes sense to compare the expected effects for the schedules which are, in some clearly defined sense, optimal for each of the populations under consideration; this represents the usual way of "standardization through optimization". In doing so, we compare maxima of potential benefits that could be gained in each setting.

The problem of screening optimization is of great interest in itself. Because of the significant cancer incidence and progress of tumor detection technology, cancer surveillance and screening are becoming increasingly important and costly public health problems. It is clear that appropriate mathematical methods are indispensable for a more effective management of the caseload through designing optimal surveillance strategies. Interest in exploring this avenue has quickened in the past few decades [16, 25-39].

The problem of optimal cancer surveillance can be set up as a search for optimal scheduling of screens subject to certain constraints on the number and timing of medical examinations. Yakovlev and Tsodikov [16] have developed methods for constructing optimal surveillance strategies based on the minimum delay time criterion, given that the total number of examinations is fixed, see also [39]. They used dynamic programming methodology to solve the associated optimization problem. However, there are two weak points in their approach. First, the probability of tumor detection is assumed to be independent of the process of tumor regrowth. Second, estimation of the tumor onset time distribution is feasible only if a sample of diagnostic times produced by a discrete surveillance program with known false negative rate is available. The same applies equally to pre-diagnosis screening programs.

To surmount the above mentioned problems we have explored a new avenue in the problem of cancer screening optimization. The newly developed methodology is well adapted to the structure of the data amassed in the Utah Population Data Base and the Utah Cancer Registry. As a measure of the effect of screening, we propose to use the difference between the expected tumor sizes at detection with and without screening, which coincides with the Kantorovich distance between the distributions of the corresponding random variables. The structure of this distance allows for characterizing the net effect of screening, as compared to that of spontaneous detection. Taken alone, the design of optimal schedules of cancer surveillance does not require information on tumor size at detection; such schedules can be constructed once the basic model parameters have been estimated in one way or another from epidemiological data at hand.

The proposed approach offers the following distinct advantages:

1. It provides a simple but still realistic description of cancer latency;
2. It can be generalized in various ways while retaining its basic structure;

3. It furnishes a biologically meaningful interpretation of data analyses;
4. It accommodates standard population-based statistical data; its implementation does not depend heavily on availability of the data yielded by screening trials;
5. Rigorous statistical methods are available for estimation of model parameters;
6. It can be used for designing optimal strategies of cancer screening and surveillance.

3.2. The Model

In describing the natural history of cancer, the process of tumor development can be broken down into three stages. These stages are:

- formation of initiated cells;
- promotion of initiated cells resulting in appearance of the first malignant clonogenic cell;
- subsequent growth and progression of malignant tumor.

The duration of each stage of carcinogenesis is thought of as a random variable. In our sample calculations, we used a two-parameter gamma family to specify the distribution of the length of the first two stages of carcinogenesis. However, more elaborate mechanistic models of carcinogenesis are available to describe the time to the event of malignant transformation at the cellular level (see Section 2.1). In particular, we plan to use the Moolgavkar-Venzon-Knudson model and the Yakovlev-Polig model in our future research.

We proceed from the following general functional form for the tumor size (the number of cells in a tumor) S :

$$S(w) = f_\theta(w),$$

where w is the time from the moment of the onset of cancer, and θ is a parameter which may be scalar or vector, deterministic or random. It is assumed that, for every θ, f_θ is a strictly monotonically increasing absolutely continuous function such that $f_\theta(0) = 1$. For a given θ, denote by g_θ the inverse function for f_θ, and set

$$\Phi_\theta(w) := \int_0^w f_\theta(u)du.$$

Specific laws of tumor growth of primary interest are:

(1) Deterministic exponential growth; in this case, $S(w) = e^{\lambda w}$, where $\lambda > 0$ is a constant growth rate, see [40] for substantiation;

(2) Exponential growth with λ thought of as a gamma distributed r. v. [41];

(3) The Gompertz law:

$$S(w) = e^{A(1-e^{-Bw})},$$

with constant parameters $A, B > 0$.

The sequence of moments of time assigned for medical exams for a specific cancer and counted from the birth of a patient will be called a *screening schedule.* Let $\mathcal{T}$ be the set of all possible screening schedules $\tau = \{\tau_1 < \tau_2 < ... < \tau_n\}$. The set $\mathcal{T}$ may be subject to (some of) the following restrictions:

(a) $n \leq n_0$, where n_0 is an upper bound for the number of exams;

(b) $\tau_1 \geq m$ and $\tau_n \leq M$, where m and M are the earliest and the latest times for the first and the last exams, respectively;

(c) $\tau_{i+1} - \tau_i \geq h > 0$ for all $i = 1, 2, ..., n-1$. This condition suggests a lower bound h for the minimal duration between any two successive exams.

Other restrictions on the moments of exams can also be accommodated. In the language of control theory, the set $\mathcal{T}$ is referred to as the set of admissible schedules.

3.3. The Screening Efficiency Functional

Numerous attempts have been made to relate the probability of detecting a tumor to its size [40-46]. Following Brown et al. [44], we assume that the rate of tumor detection is proportional to the current tumor size.

We distinguish between *spontaneous* and *screening based* tumor detections. The first occurs in the absence of or concurrently with screening and is thought of as a continuous process. In contrast to this, screening based detection is an instantaneous event that may occur only at the moments of the prescribed medical exams and is therefore a discrete process. When both types of detection are present, they can be viewed as competing risks.

Let random variables W_0 and W_1, denote the times of spontaneous and screening based detections, measured from the moment of cancer onset, respectively, and let T be the time of tumor onset. We have derived a formula for the screening efficiency functional proceeding from the following two biologically natural assumptions.

1. The r.v.'s W_0 and T are independent.

2. For every $t \geq 0$, the r.v.'s W_1 and W_0 are conditionally independent given that $T = t$.

The first assumption implies that the moment of spontaneous tumor detection measured from the appearance of the first malignant clonogenic cell is independent of the prior duration of tumor latency. The second assumption reflects a technological (or instrumental) nature of both detection processes. It states that, given the moment of cancer onset, the two times W_0 and W_1, at which competing events of the spontaneous and screening based tumor detection may occur, are independent. This statement immediately follows from the assumption that both detection processes are completely determined by the current tumor size as a deterministic function of time.

For an admissible screening schedule $\tau \in \mathcal{T}$, we define the efficiency functional as the Kantorovich distance $d_K(N_0, N; \tau)$ between the tumor sizes N_0 and N at spontaneous and combined detection. It is well known [47, 48] that

$$d(N, N_0; \tau) = \int_1^\infty \mid \bar{F}_{N_0}(n) - \bar{F}_N(n) \mid dn.$$

An alternative expression for the efficiency functional is given by

$$d(N, N_0; \tau) = \int_1^\infty \bar{F}_{N_0}(n)dn - \int_1^\infty \bar{F}_N(n)dn = EN_0 - EN,$$

where E stands for the expectation.

An explicit analytic expression for screening efficiency functional is presented in the attached paper by Hanin et al. (accepted for publication in *Mathematical and Computer Modelling*).

The optimization problem

$$d(N, N_0; \tau) \to \max, \quad \tau \in \mathcal{T},$$

has been solved by exhaustive search with some simplification arising from the special form of the dependence of the efficiency functional on τ.

3.4. Numerical Experiments

The purpose of our numerical experiments was to check feasibility of numerical and optimization problems associated with the proposed approach to stochastic modeling of cancer screening. These experiments are described at length in the attached paper by Hanin et al. The most interesting finding was that in the case of exponential tumor growth optimal screening schedules are uniform or very close to such. The parameter values used in our sample computations were judiciously chosen on the basis of our experience; they are no better than an educated guess. However, this study clearly shows that mathematical and computational problems of optimal cancer surveillance are tractable within the framework of the proposed model. What is now required is to make a final stride towards the use of parameter estimates obtained from population-based epidemiologic data.

3.5. Future Plans

The model will be validated using the relevant epidemiological data on birth cohorts identified through the Utah Population Data Base. An appealing possibility would be to estimate unknown parameters of breast cancer latency solely from the age-at-diagnosis data. However, this is unlikely to be feasible given the number of parameters incorporated into the model. The main obstacle is that the available information is too sparse for estimating all the parameters, and a search for additional sources of data is clearly warranted. On the other hand, our modeling techniques allow incorporation of the process of tumor detection into mechanistic models of tumor latency, thereby making it possible to utilize data on tumor size at detection as an additional source of information on the natural history of the disease. This information is available in the Utah Cancer Registry. For a given functional form of tumor growth, we will obtain within the framework of our model the corresponding joint distribution of tumor size and age at detection. Proceeding from this joint distribution we will construct the likelihood of the sample under study. The likelihood function will be maximized numerically by computer using nonlinear

programming methods [24]. Once the estimation problem has been solved we are in a position to design optimal schedules of breast cancer surveillance allowing for individual information on family history.

Key Research Accomplishments

Our key accomplishments in Year 1 can be summarized briefly as follows:

• We have developed computer programs implementing four statistical procedures for estimation of the hazard function; these procedures accommodate data subjected to random truncation and censoring.

• A new method has been developed for designing optimal schedules of breast cancer surveillance specially adapted to population-based settings.

• Numerical experiments have shown that mathematical and computational problems of optimal cancer surveillance are tractable within the framework of the proposed model of cancer surveillance and detection.

Reportable Outcomes

1. Hanin, L.G., Tsodikov, A.D., and Yakovlev, A.Y. Optimal schedules of cancer surveillance and tumor size at detection, *Mathematical and Computer Modelling*, in press (see Appendix).
2. Hanin, L.G. Optimal schedules of cancer surveillance, presented at the International Workshop on Cure Rate Estimation, Tampa, Florida, February 19-21, 1999.
3. Yakovlev, A.Y., Tsodikov, A.D., and Hanin, L.G. Optimal schedules of breast cancer surveillance, Abstract, Era of Hope Meeting, Atlanta, June 2000.
4. Boucher, K.M. and Kerber, R.A. The shape of the hazard function for cancer incidence, Abstract, Era of Hope Meeting, Atlanta, June 2000.
5. Yakovlev, A.Y. Mechanistic Modeling of Breast Cancer Surveillance, Grant Application, RFA "Cancer Intervention and Surveillance Network (CISNET)", NIH/NCI.

Conclusions

We have developed a mathematical model yielding an algorithm for designing optimal schedules of breast cancer surveillance. An explicit expression of the screening efficiency functional has been derived. The derivation is based on a plausible assumption that the intensity of detection (the hazard function for the age at detection) is proportional to the current tumor size. The main advantage of the proposed approach is that it accommodates cohort data of a fairly general structure, not only the data resulting from screening trials. We have also developed several numerical algorithms and software for estimating the hazard function for breast cancer incidence from the data amassed in the Utah Population Data Base. Allowing for the effects of random censoring and truncation, these procedures will be used in future research

for testing covariate effects associated with different indicators of family history.

So what? We now have the necessary tools for:

- designing optimal schedules of breast cancer surveillance given the numerical parameters describing the natural history of the disease are known;
- testing covariate effects associated with different indicators designed to aggregate family history information.

In Year 2, our focus will be on the development of methods for estimation of the natural history of breast cancer from the data available in the Utah Population Data Base and the Utah Cancer Registry. Using these resources we will also test indicators of family history in order to select the most informative one for the purposes of individualization of breast cancer surveillance strategies.

References

1. A.Y. Yakovlev, A.D. Tsodikov, K. Boucher, and R. Kerber, The shape of the hazard function for breast cancer: curability of the disease revisited. *Cancer* **85**, 1789-1798, (1999).

2. O. O. Aalen, Modeling heterogeneity in survival analysis by the compound Poisson distribution. *Annals of Applied Probability* **2**, 951-972, (1992).

3. D. F. Easton, The inherited component of cancer. *British Medical Bulliten* **50**, 527-535, (1994).

4. J. P. Struewing, P. Hartge, S. Wacholder, S. M. Baker, M. Berlin, M. McAdams, M. M. Timmerman, L. C. Brody, and M. A. Tucker, The risk of cancer associated with specific mutations of BRCA1 and BRCA2 among Ashkenazi Jews. *N. Engl. J. Med.* **336**, 1401-1408, (1997).

5. D. Ford and D. F.Easton, The genetics of breast and ovarian cancer. *Br. J. Cancer* **72**, 805-812, (1995).

6. B. Newman, H. Mu, L. M. Butler, R. C. Millikan, P. G. Moorman, and M.C. King, Frequency of breast cancer attributable to BRCA1 in a population-based series of American women *JAMA* **279**, 915-921, (1998).

7. T. I . Anderson, Genetic heterogeneity in breast cancer susceptibility. *Acta Oncol.* **35**, 407-410, (1996).

8. P. Vehmanen, L.S. Friedman, H. Eerola, L. Sarantaus, S. Pyrhönen, B.A.J. Ponder, T. Muhonen et al., A low proportion of BRCA2 mutations in Finnish breast cancer families. *Am J. Human Genet.* **60**, 1050-1058, (1997).

9. S. H. Moolgavkar and D. J. Venzon, Two-event models for carcinogenesis: incidence curves for childhood and adult tumors. *Math. Biosci.* **47**, 55-77, (1979).

10. S. H. Moolgavkar and A.Knudson, Mutation and cancer: a model for human carcinogenesis. *J. Natl Cancer Institute* **66**, 1037-1052, (1981).

11. S. H. Moolgavkar and E. G. Luebeck, Two-event model for carcinogenesis: biological, mathematical and statistical considerations. *Risk Anal.* **10**, 323-341, (1990).

12. W. F. Heidenreich, E. G. Luebeck and S. H. Moolgavkar, Some properties of the hazard function of the two-mutation clonal expansion model. *Risk Analysis* **17**, 391-399, (1997).

13. A. Kopp-Schneider, C.J. Portier, and C. D. Sherman, The exact formula for tumor incidence in the two-stage model. *Risk Analysis* **14**, 1079-1080, (1994).

14. Q. Zheng, On the exact hazard and survival functions of the MVK stochastic carcinogenesis model. *Risk Anal.* **14**, 1081-1084, (1994).

15. L. G. Hanin and A. Y. Yakovlev, A nonidentifiability aspect of the the two-stage model of carcinogenesis. *Risk Anal.* **16**, 711-715, (1996).

16. A. Yu. Yakovlev and A. D. Tsodikov, *Stochastic Models of Tumor Latency and Their Biostatistical Applications*, World Scientific, Singapore (1996).

17. N. Keiding Independent delayed entry. In *Survival Analysis: the State of the Art*, J. P. Klein and P. K. Goel, eds., Kluwer, Boston-Dordrecht-London, 1992, pp. 309-326.

18. M. Woodroofe, Estimating a distribution function with truncated data. *Ann. Statist.* **13**, 163-177, (1985).

19. E. L. Kaplan and P. Meier, Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**, 457-481, (1958).

20. N. Keiding and R. D. Gill, Random trucation models and Markov processes *Ann. Statist.* **18**, 582-602, (1990).

21. S. Johansen, The product limit as a maximum likelihood estimator. *Scand J. Statist.* **5**, 195-199, (1978).

22. M.-C. Wang, N. P. Jewell, and W.-Y. Tsai, Asymptotic properties of the product limit estimate under random truncation *Ann. Statist.* **14**, 1597-1605, (1986).

23. J. Etezadi-Amoli and A. Ciampi, Extended hazard regression for censored survival data with covariates: a spline approximation for the baseline hazard function. *Biometrics* **43**, 181-192, (1987).

24. D.M. Himmelblau, *Applied Nonlinear Programming* McGraw-Hill, Austin, 1972.

25. R. Kirch and M. Klein, Surveillance schedules for medical examinations, *Management Science* **20**, 1403-1409 (1974).

26. A.K. Shahani and D.M. Crease, Towards models of screening for early detection of disease, *Adv. Appl. Prob.* **9**, 665-680 (1977).

27. D.M. Eddy, A mathematical model for timing repeated medical tests, *Medical Decision Making* **3**, 34-62 (1983).

28. A.D. Tsodikov and A.Yu. Yakovlev, On the optimal policies of cancer screening, *Math. Biosci.* **107**, 21-45 (1991).

29. A.D. Tsodikov, A.Yu. Yakovlev and L. Petukhov, Some approaches to screening optimization. In *Statistique des Processus en Milieu Médical* (edited by B. Bru, C. Huber and B. Prum) Université Paris V, Paris, France, 1-48 (1991).

30. A.D. Tsodikov, Screening under uncertainty. Games approach, *Syst. Anal. Model. Simul.* **9**, 259-262 (1992).

31. G. Parmigiani, On optimal screening schedules, *Biometric Bulletin* **8, 3**, 21 (1991).

32. G. Parmigiani, Optimal scheduling of fallible inspections, *DP 92-A38*, ISDS, Duke University (1992).

33. G. Parmigiani, On optimal screening ages, *J. Amer. Statist. Assoc.* **88**, 622-628 (1993).

34. G. Parmigiani, Timing medical examinations via intensity functions, *Biometrika* **84, 4**, 803-816 (1997).

35. G. Parmigiani and M.S. Kamlet, A cost-utility analysis of alternative strategies in screening for breast cancer. In *Bayesian Statistics in Science and Technology: Case Studies* (edited by C. Gatsonis, J. Hodges, R.E. Kass and N. Singpurwalla), Springer, New York, 390-402 (1993).

36. M. Zelen, Optimal scheduling of examinations for the early detection of disease, *Biometrika* **80**, 279-293 (1993).

37. Ping Hu and M. Zelen, Planning clinical trials to evaluate early detection programmes, *Biometrika* **84, 4**, 817-829 (1997).

38. S.J. Lee and M. Zelen, Scheduling periodic examinations for the early detection of disease: applications to breast cancer, *J. Amer. Statist. Ass.* **93**, 1271-1281 (1998).

39. A.D. Tsodikov, B. Asselain, A. Fourquet, T. Hoang and A.Yu. Yakovlev, Discrete strategies of cancer post-treatment surveillance. Estimation and optimization problems, *Biometrics* **51**, 437-447 (1995).

40. R. Bartoszyński, A modeling approach to metastatic progression of cancer. In *Cancer Modeling* (edited by J.R. Thompson and B.W. Brown), Marcel Dekker, New York and Basel, 237-267 (1987).

41. M. Klein and R. Bartoszyński, Estimation of growth and metastatic rates of primary breast cancer. In: *Mathematical Population Dynamics* (edited by O. Arino, D. E. Axelrod and M. Kimmel), Marcel Dekker, New York, 397-412 (1991).

42. N.E. Atkinson, R. Bartoszyński, B.W. Brown and J.R. Thompson, On estimating the growth function of tumors, *Math. Biosci.* **67**, 145-166 (1983).

43. N.E. Atkinson, B.W. Brown and J.R. Thompson, On the lack of concordance between primary and secondary tumor growth rates, *J. Natl. Cancer Inst.* **78**, 425-435 (1987).

44. B.W. Brown, N.E. Atkinson, R. Bartoszyński and E.D. Montague, Estimation of human tumor growth rate from distribution of tumor size at detection, *J. Natl. Cancer Inst.* **72**, 31-38 (1984).

45. A.Yu. Yakovlev, L.G. Hanin, S.T.. Rachev and A.D. Tsodikov, Distribution of tumor size at detection and its limiting form, *Proc. Natl. Acad. Sci. USA* **93**, 6671-6675 (1996).

46. L.G. Hanin, S.T. Rachev, A.D. Tsodikov and A.Yu. Yakovlev, A stochastic model of carcinogenesis and tumor size at detection, *Adv. Appl. Prob.* **29**, 607-628 (1997).

47. L.V. Kantorovich and G.P. Akilov, *Functional Analysis*, 2nd ed., Pergamon Press, New York (1982).

48. S.T. Rachev and R.M. Short, Duality theorems for Kantorovich-Rubinstein and Wasserstein functionals, *Dissertationes Math.* **299** (1990).

Appendix

Optimal Schedules of Cancer Surveillance and Tumor Size at Detection

L. G. HANIN

Department of Mathematics, Idaho State University, Pocatello, ID 83209-8085, U.S.A.

A. D. TSODIKOV and A. YU. YAKOVLEV

Huntsman Cancer Institute and Department of Oncological Sciences, University of Utah, 2000 East North Campus Drive, Salt Lake City, UT 84112-5550, U.S.A.

Dedicated to the memory of Robert Bartoszyński

ABSTRACT – The paper explores methodological and mathematical aspects of a new approach to constructing optimal schedules of cancer screening. This approach consists in systematic use of tumor size at detection, combining stochastic models of tumor latency, tumor growth and tumor detection, and employing a new biologically natural screening efficiency criterion defined as the Kantorovich distance between the tumor size at spontaneous detection in the absence of screening and the tumor size at detection when both spontaneous and screening based mechanisms are in place. An explicit formula for the efficiency functional is obtained. Sample calculations suggest that in the case of exponential tumor growth the optimal screening schedules with a fixed number of exams have a trend to uniformity.

Keywords – Screening, Optimal schedules, Tumor onset, Tumor size, Carcinogenesis models

1. INTRODUCTION

Because of the significant cancer incidence and progress of tumor detection technology, cancer surveillance and screening are becoming increasingly important and costly public health problems. It is clear that appropriate mathematical methods are indispensable for a more effective management of the caseload through designing optimal surveillance strategies. Interest in exploring this avenue has quickened in the past few years [1-17].

The present work discusses methodological aspects of a new approach to optimization of cancer screening allowing for cancer detection at the earliest stages of tumor development. This makes the chances of tumor cure more favorable, reducing

the probability of tumor recurrence. The problem of optimal cancer surveillance is set up as a search for optimal scheduling of screens subject to certain constraints on the number and timing of medical exams. Problems of a similar nature have already been addressed in the literature. Yakovlev and Tsodikov [1] have developed methods for constructing optimal surveillance strategies based on the minimum delay time criterion, given that the total number of examinations is fixed, see also [2]. They used dynamic programming methodology to solve the associated optimization problem. Their results show that this approach holds much promise for further practical use. As one example, the current practice for the breast cancer post-treatment surveillance at Curie Institute (Paris, France) is to examine the patients once per semester for the first 4 years, once per year for the next 6 years, and once every 2 years for the remaining period. For this strategy, the estimated false negative rate appears to be equal to 0.2 with the mean delay of the recurrence detection 4.1 months. Taking advantage of a previously proposed parametric model of tumor recurrence [3], the authors constructed the optimal strategy that provides a 33% reduction in the delay time, with the tests that comprise the optimal surveillance schedule tending to be more frequent when the hazard rate for the time to tumor is high. However, there are two weak points in this approach. First, the probability of tumor detection is assumed to be independent of the process of tumor regrowth. Second, estimation of the tumor onset time distribution is feasible only if a sample of diagnostic times produced by a discrete surveillance program with known false negative rate is available. The same applies equally to pre-diagnosis screening programs.

An alternative approach to the problem is to minimize the average cost of surveillance accounting for both examination costs and costs of late detection [1], [4-15]. Since the two cost constituents are linked in the optimization procedure, the cost-utility approach makes it possible to search for both the optimal number of examinations and their sequence in time. However, the costs of late detection are usually very difficult to evaluate. For yet another optimization criterion based on the power of a statistical test for mortality rates, the reader is referred to [16].

Focusing our effort on possible medical rather than economic benefits, we propose to explore a new approach to the problem which is based on tumor size at detection. Tumor size is one of the most clinically significant characteristics of tumor maturity that determines largely the probability of both spontaneous and screening based tumor detection. This approach makes it possible to utilize data on tumor size at detection as an additional source of information on the natural history of the disease; some readily available epidemiologic data obtained from the control population in the absence of screening appear to be sufficient for estimation purposes. Another advantage of this approach is that it offers a natural way for incorporating the stage of tumor progression, where cancer detection normally occurs, into stochastic models of carcinogenesis. The proposed model of tumor progression accommodates a wide range of deterministic and stochastic laws of tumor growth.

As a measure of the effect of screening, we propose to use the difference between the expected tumor sizes at detection with and without screening, which coincides with the Kantorovich distance [18-21] between the distributions of the corresponding random variables. The structure of this distance allows for characterizing the net

effect of screening, as compared to that of spontaneous detection.

Further advancements of the proposed approach to constructing optimal schedules of cancer screening will hopefully give answers to the following questions of major theoretical and practical importance:

1. Is the optimal efficiency of screening high enough to warrant its implementation?

2. What is the relation between the optimal screening schedules and their efficiencies for the criteria based on the tumor size and the expected time delay?

3. What are cancer specific patterns of optimal screening schedules?

4. What is the impact of hypothesized laws of tumor growth on the optimal screening efficiency and the pattern of the optimal examination schedules?

5. What are quantitative characteristics of the initiation, promotion, and progression stages for specific cancers?

The structure of the present paper is as follows. In Section 2, we describe some models of the natural history of cancer (including cancer latency and growth), screening schedules, and cancer detection. Here, we also formulate basic assumptions and introduce mathematical formalism. An explicit formula for the efficiency functional is derived in Section 3. Sample numerical calculations and analysis of their results are addressed in Section 4.

2. BASIC NOTIONS

2.1. Models of carcinogenesis

In describing the natural history of cancer, the process of tumor development can be broken down into three stages. These stages are:

- formation of initiated cells;
- promotion of initiated cells resulting in appearance of the first malignant clonogenic cell;
- subsequent growth and progression of malignant tumor.

The duration of each stage of carcinogenesis is thought of as a random variable (r.v.). In our sample calculations presented in Section 4, we use a two-parameter gamma family to specify the distribution of the length of the first two stages of carcinogenesis. However, more elaborate mechanistic models of carcinogenesis are available to describe the time to the event of malignant transformation. We provide two examples of such models.

The most widely accepted model of tumor latency is commonly referred to as the Moolgavkar-Venzon-Knudson (MVK) model [22, 23]. This Markovian two-stage model involves four parameters that refer to the rates of initiation of target stem cells (that is, formation of primary precancerous lesions), and rates of division, death or differentiation, and malignant transformation of initiated cells. It was first pointed out by Heidenreich [24] and subsequently by Hanin and Yakovlev [25] and Heidenreich, Luebeck and Moolgavkar [26] that these four parameters are not jointly identifiable from time-to-tumor data. In the case of constant parameters, all triples of their

identifiable combinations were described at length in [25]. In the latter case, the MVK model leads to the following explicit formula for the distribution of the total duration T of the first two stages, that is, of the time from the birth of an individual to the tumor onset [27, 28]:

$$\bar{F}_T(t) := Pr(T > t) = \left[\frac{(a+b)e^{at}}{b + ae^{(a+b)t}}\right]^{\rho}, \quad t \geq 0. \tag{1}$$

Here $a,\ b,\ \rho > 0$ are identifiable parameters of the model, $\bar{F}_T := 1 - F_T$ is the survivor function of the r.v. T, and F_T is the cumulative distribution function (c.d.f.) of the r.v. T.

Another model of carcinogenesis was proposed by Yakovlev and Polig in [29]. According to this model, the hazard function ϕ of the time T of tumor latency, which is related to the survivor function by

$$\bar{F}_T(t) = e^{-\int_0^t \phi(s)ds}, \quad t \geq 0, \tag{2}$$

is of the form

$$\phi(s) = \theta_1 e^{-\theta_2 \int_0^s h(u)du} \int_0^s h(u) f(s-u)du, \quad s \geq 0, \tag{3}$$

where h is a given time-dependent rate of external exposure, f is the probability density function (p.d.f.) of the tumor promotion time, and $\theta_1,\ \theta_2$ are positive constants. The key feature of the Yakovlev-Polig model is that it allows for the process of cell death to compete with the process of tumor promotion. Two particular cases of the model referring to spontaneous and induced carcinogenesis were employed in [30] and [31] to study the distribution of tumor size under a threshold type mechanism of tumor detection. Recently, Hanin and Boucher [32] found conditions under which the parameters f, θ_1, θ_2 of the model given by (3) are identifiable from time-to-tumor observations. Specifically, a general necessary condition for identifiability of model (3) is given by the following theorem.

Theorem 1. *Suppose that the function h satisfies $\int_0^\infty h(t)dt < \infty$ and that, for some $C > 0$, $h(t) = 0$ for $t > C$. If the model is identifiable in a family $\mathcal{F}$ then*

$$F(C) > 0 \quad \text{for all} \quad F \in \mathcal{F}.$$

Definition. A family $\mathcal{F}$ of absolutely continuous probability distributions on $\mathbf{R}_+$ is said to be *graduated* if for every two distinct p.d.f's $f,\ \tilde{f} \in \mathcal{F}$ and for every constant $A > 0$, there is a number $\tau > 0$ (which may depend on $f,\ \tilde{f}$, and A) such that either $Af(t) > \tilde{f}(t)$ for all $t \geq \tau$ or $Af(t) < \tilde{f}(t)$ for all $t \geq \tau$.

The following result generalizes Theorem 1 in the case of graduated families.

Theorem 2. *Suppose that h is bounded, supported on $[0, C]$ for some $C > 0$, and positive almost everywhere on $[0, C]$. Then the model is identifiable in a graduated family $\mathcal{F}$ if and only if $F(C) > 0$ for all $F \in \mathcal{F}$.*

2.2. Tumor growth

The following general functional form is assumed for the tumor size (the number of cells in a tumor) S :

$$S(w) = f_\theta(w), \tag{4}$$

where w is the time from the moment of the onset of cancer, and θ is a parameter which may be scalar or vector, deterministic or random. It is assumed that, for every θ, f_θ is a strictly monotonically increasing absolutely continuous function such that $f_\theta(0) = 1$. For a given θ, denote by g_θ the inverse function for f_θ, and set

$$\Phi_\theta(w) := \int_0^w f_\theta(u)du.$$

Specific laws of tumor growth of primary interest are:

(1) Deterministic exponential growth; in this case, $S(w) = e^{\lambda w}$, where $\lambda > 0$ is a constant growth rate, see [33] for substantiation;

(2) Exponential growth with λ thought of as a gamma distributed r. v. [34];

(3) The Gompertz law:

$$S(w) = e^{A(1-e^{-Bw})},$$

with constant parameters $A, B > 0$.

2.3. Screening schedules

The sequence of moments of time assigned for medical exams for a specific cancer and counted from the birth of a patient will be called a *screening schedule*. Let $\mathcal{T}$ be the set of all possible screening schedules $\tau = \{\tau_1 < \tau_2 < ... < \tau_n\}$. The set $\mathcal{T}$ may be subject to (some of) the following restrictions:

(a) $n \leq n_0$, where n_0 is an upper bound for the number of exams;

(b) $\tau_1 \geq m$ and $\tau_n \leq M$, where m and M are the earliest and the latest times for the first and the last exams, respectively;

(c) $\tau_{i+1} - \tau_i \geq h > 0$ for all $i = 1, 2, ..., n-1$. This condition suggests a lower bound h for the minimal duration between any two successive exams.

Other restrictions on the moments of exams can also be accommodated. In the language of control theory, the set $\mathcal{T}$ is referred to as the set of admissible schedules.

2.4. Tumor detection

We distinguish between *spontaneous* and *screening based* tumor detections. The first occurs in the absence of or concurrently with screening and is thought of as a continuous process. In contrast to this, screening based detection is an instantaneous event that may occur only at the moments of the prescribed medical exams and is therefore a discrete process. When both types of detection are present, they can be viewed as competing risks.

Numerous attempts have been made to relate the probability of detecting a tumor to its size [33-37]. Following Brown et al. [37], we assume that the rate r_0 of spontaneous tumor detection is proportional to the current tumor size:

$$r_0 = \alpha_0 S, \tag{5}$$

where α_0 is a positive constant.

Let r.v.'s W_0 and W_1 denote the times of spontaneous and screening based detections, counted from the moment of cancer onset, respectively. Then for the moment W of combined detection, when both detection mechanisms are in place, we have $W = min(W_0, W_1)$. Denote by

$$N_0 = f_\theta(W_0) \quad \text{and} \quad N = f_\theta(W) \tag{6}$$

the corresponding tumor sizes at spontaneous and combined detection.

Keeping in mind relation (2) between the survivor function of an absolutely continuous nonnegative r.v. and its hazard rate, we derive from (5) that, in the case of non-random parameter θ,

$$\bar{F}_{W_0}(w) = e^{-\int_0^w r_0(u)du} = e^{-\alpha_0 \int_0^w f_\theta(u)du} = e^{-\alpha_0 \Phi_\theta(w)}. \tag{7}$$

Therefore,

$$\bar{F}_{N_0}(n) = \bar{F}_{W_0}(g_\theta(n)) = e^{-\alpha_0 \Phi_\theta(g_\theta(n))},$$

and hence

$$EN_0 = 1 + \int_1^\infty \bar{F}_{N_0}(n)dn = 1 + \int_1^\infty e^{-\alpha_0 \Phi_\theta(g_\theta(n))}dn = 1 + \int_0^\infty e^{-\alpha_0 \Phi_\theta(u)} f'_\theta(u)du. \tag{8}$$

If θ is a r.v. then an additional integration in (8) with respect to the distribution of θ is required.

In particular, for non-random exponential tumor growth with rate λ, we have

$$\bar{F}_{W_0}(w) = e^{-\frac{\alpha_0}{\lambda}(e^{\lambda w}-1)}, \quad w \geq 0, \tag{9}$$

$$\bar{F}_{N_0}(n) = e^{-\frac{\alpha_0}{\lambda}(n-1)}, \quad n \geq 1, \tag{10}$$

and

$$EN_0 = 1 + \frac{\lambda}{\alpha_0}. \tag{11}$$

Equation (10) suggests that in this case the r.v. N_0 has a translated exponential distribution with parameter α_0/λ. If λ is a r.v. which is gamma distributed with parameters μ, ν, then it follows from (11) that

$$EN_0 = 1 + \frac{\mu}{\alpha_0 \nu}.$$

We now specify the distribution of the r.v. W_1. Recall that W_1 is the time of screening based detection (in the absence of spontaneous detection) counted from the moment of appearance of the first malignant clonogenic cell. Indeed, the distribution of W_1 depends on the selected screening schedule $\tau = \{\tau_1 < \tau_2 < ... < \tau_n\}$. For the sake of convenience, set $\tau_0 := 0$ and $\tau_{n+1} := \infty$. It suffices to define, for every $t \geq 0$, the conditional distribution of W_1 given that $T = t$.

Let $\tau_i \leq t < \tau_{i+1}$, $0 \leq i \leq n$. For $0 \leq i \leq n-1$ and $i+1 \leq k \leq n$, define the probability $p_t(k) := Pr(W_1 = \tau_k - t | T = t)$ of tumor detection at the $k-$th

screen given the cancer onset at moment t, and by $p_t(\infty) = 1 - \sum_{k=i+1}^{n} p_t(k)$ the corresponding conditional probability that tumor is not detected by screening.

We introduce a discrete analogue of the hazard rate for the screening based detection by

$$\mu_t = \sum_{k=i+1}^{n} r_t(k)\delta_{\tau_k - t}, \tag{12}$$

where δ_x stands for the Dirac measure at x, and the sum over the empty set of indices is set, as usual, to be zero. By definition, the discrete measure μ_t is related to the conditional survivor function of W_1 given that $T = t$ through the equation

$$\bar{F}_{W_1|T=t}(w) = e^{-\int_0^w d\mu_t(u)}, \quad w \geq 0, \tag{13}$$

compare with (2). It follows from (12) and (13) that

$$\sum_{j=k+1}^{n} p_t(j) + p_t(\infty) = [\sum_{j=k}^{n} p_t(j) + p_t(\infty)]e^{-r_t(k)}$$

or, equivalently, that

$$1 - \sum_{j=i+1}^{k} p_t(j) = [1 - \sum_{j=i+1}^{k-1} p_t(j)]e^{-r_t(k)}. \tag{14}$$

For $k = i + 1$, we find from (14) that

$$1 - p_t(i+1) = e^{-r_t(i+1)}. \tag{15}$$

More generally, iterating this argument we obtain that

$$p_t(k) = e^{-\sum_{j=i+1}^{k-1} r_t(j)}[1 - e^{-r_t(k)}], \quad i + 1 \leq k \leq n.$$

Observe that this holds true for all $k = 1, ..., n$, if we set $p_t(k) = r_t(k) = 0$ for $1 \leq k \leq i$.

Similar to (5), we are assuming that the discrete rate of screening based detection is proportional to the current tumor size:

$$r_t(k) = \alpha S(\tau_k - t), \quad i + 1 \leq k \leq n, \tag{16}$$

with some constant $\alpha > 0$. Combining (13), (12) and (16) with (4) we find that, given any t such that $\tau_i \leq t < \tau_{i+1}$, $0 \leq i \leq n - 1$,

$$\bar{F}_{W_1|T=t}(w) = e^{-\alpha \sum_{k=i+1}^{j} f_\theta(\tau_k - t)}, \quad \text{where} \quad \tau_j - t \leq w < \tau_{j+1} - t, \quad i + 1 \leq j \leq n. \tag{17}$$

Consider the case of one exam occurring at a moment τ with the detection probability $p = p(t, \tau)$ and the discrete detection rate $r = r(t, \tau)$. Then by (15), $1 - p = e^{-r}$. If the probability p is small then the rate r is approximately equal to p. In particular, under the assumtion (16), the probability of tumor detection is approximately proportional to the current tumor size: $p \simeq \alpha S(\tau - t)$. Klein and Bartoszyński [34] proceeded in their study of breast cancer from a more general assumption that the probability of tumor detection is proportional to some power of the tumor size. Their estimate of this power leads, however, to a value which is very close to 1.

3. FORMULA FOR THE SCREENING EFFICIENCY FUNCTIONAL

We proceed from the following two biologically natural assumptions.

1. The r.v.'s W_0 and T are independent.

2. For every $t \geq 0$, the r.v.'s W_1 and W_0 are conditionally independent given that $T = t$.

The first assumption claims that the moment of spontaneous tumor detection measured from the appearance of the first malignant clonogenic cell is independent of the prior duration of tumor latency. The second assumption reflects a technological (or instrumental) nature of both detection processes. It states that, given the moment of cancer onset, the two times W_0 and W_1, at which competing events of the spontaneous and screening based tumor detection may occur, are independent. This statement immediately follows from the assumption that both detection processes are completely determined by the current tumor size as a deterministic function of time.

For an admissible screening schedule $\tau \in \mathcal{T}$, we define the efficiency functional as the Kantorovich distance $d_K(N_0, N; \tau)$ (see [18], [20], [21]) between the tumor sizes N_0 and N at spontaneous and combined detection. This quantity serves as a clinically natural measure of the gain resulting from screening. It is well known [19], [20] that

$$d(N, N_0; \tau) = \int_1^\infty \mid \bar{F}_{N_0}(n) - \bar{F}_N(n) \mid dn. \tag{18}$$

It follows from (7), inequality $W_0 \geq W$, and monotonicity of the function f_θ that the r.v. N_0 stochastically dominates the r.v. N : $\bar{F}_{N_0} \geq \bar{F}_N$. This leads to the following alternative expression for the efficiency functional:

$$d(N, N_0; \tau) = \int_1^\infty \bar{F}_{N_0}(n)dn - \int_1^\infty \bar{F}_N(n)dn = EN_0 - EN, \tag{19}$$

where E stands for the expectation.

Suppose that parameter θ is non-random. We set $n = f_\theta(w)$ and condition upon the r.v. T in (18) to obtain

$$d(N, N_0; \tau) = \int_0^\infty \mid \bar{F}_{W_0}(w) - \bar{F}_W(w) \mid f'_\theta(w)dw$$

$$= \int_0^\infty \int_0^\infty \mid \bar{F}_{W_0}(w) - \bar{F}_{W|T=t}(w) \mid f'_\theta(w)dwdF_T(t),$$

where $\bar{F}_{W|T=t}$ is the conditional survivor function of the r.v. W given that $T = t$. Since $W = min(W_0, W_1)$, it follows from our assumptions 1 and 2 that

$$\bar{F}_{W_0} - \bar{F}_{W|T=t} = \bar{F}_{W_0} - \bar{F}_{W_0}\bar{F}_{W_1|T=t} = \bar{F}_{W_0}F_{W_1|T=t}.$$

Therefore,

$$d(N, N_0; \tau) = \int_0^\infty \int_0^\infty F_{W_1|T=t}(w)\bar{F}_{W_0}(w)f'_\theta(w)dwdF_T(t). \tag{20}$$

Observe that if $T = t$, where $\tau_i \le t < \tau_{i+1}$, $0 \le i \le n$, then the only possible values of the r.v. W_1 are $\tau_{i+1} - t, ..., \tau_{n+1} - t$. More specifically, $W_1 = \tau_j - t$, $i+1 \le j \le n$, if the $j-$th exam detected a tumor, and $W_1 = \tau_{n+1} - t = \infty$ if the tumor was not detected in the course of screening. Therefore, if $t \ge \tau_n$ or $\tau_i \le t < \tau_{i+1}$, $0 \le i \le n-1$, and $0 \le w < \tau_{i+1} - t$, then $F_{W_1|T=t}(w) = 0$. This allows us to rewrite (20) in the form

$$d(N, N_0; \tau) = \sum_{i=0}^{n-1} \int_{\tau_i}^{\tau_{i+1}} \sum_{j=i+1}^{n} \int_{\tau_j - t}^{\tau_{j+1} - t} F_{W_1|T=t}(w) \bar{F}_{W_0}(w) f_\theta'(w) dw dF_T(t).$$

We now recall the explicit expression (17) derived above for the function $\bar{F}_{W_1|T=t}$, and denote

$$G_\theta(x) := \int_x^\infty \bar{F}_{W_0}(w) f_\theta'(w) dw, \quad x \ge 0,$$

to obtain finally

$$d(N, N_0; \tau) = \sum_{i=0}^{n-1} \int_{\tau_i}^{\tau_{i+1}} \sum_{j=i+1}^{n} [1 - e^{-\alpha \sum_{k=i+1}^{j} f_\theta(\tau_k - t)}][G_\theta(\tau_j - t) - G_\theta(\tau_{j+1} - t)] dF_T(t)$$

$$= \sum_{i=0}^{n-1} \int_{\tau_i}^{\tau_{i+1}} \sum_{j=i+1}^{n} e^{-\alpha \sum_{k=i+1}^{j-1} f_\theta(\tau_k - t)} [1 - e^{-\alpha f_\theta(\tau_j - t)}] G(\tau_j - t) dF_T(t). \tag{21}$$

In the case when parameter θ is random, the right-hand side of (21) should be integrated additionally with respect to the distribution of θ.

If, in particular, $f_\theta(w) = e^{\lambda w}$ with a constant rate λ, then invoking (9) we find easily that

$$G_\theta(x) = \frac{\lambda}{\alpha_0} e^{-\frac{\alpha_0}{\lambda}(e^{\lambda x} - 1)}, \quad x \ge 0.$$

In this case the efficiency functional (21) takes on the form

$$d(N, N_0; \tau) = \frac{\lambda}{\alpha_0} \sum_{i=0}^{n-1} \int_{\tau_i}^{\tau_{i+1}} \sum_{j=i+1}^{n} e^{-\alpha \sum_{k=i+1}^{j-1} e^{\lambda(\tau_k - t)}} [1 - e^{-\alpha e^{\lambda(\tau_j - t)}}] e^{-\frac{\alpha_0}{\lambda}(e^{\lambda(\tau_j - t)} - 1)} dF_T(t). \tag{22}$$

Observe also that (19) implies

$$EN = EN_0 - d(N, N_0; \tau).$$

This allows for an explicit calculation of the expected tumor size at combined detection on the basis of formulas (8) and (21).

The problem

$$d(N, N_0; \tau) \to \max, \quad \tau \in \mathcal{T}, \tag{23}$$

can be solved by exhaustive search with some simplification arising from the special form of the dependence of the functional (21) on τ. A question of practical importance is what are the values of the number n of exams for which the problem (23) is computationally feasible. We will conclude this paper, which deals primarily with methodological and mathematical aspects of the problem of optimization of cancer

surveillance, with some sample calculations with prescribed values of model parameters.

4. NUMERICAL EXPERIMENTS

It was assumed that the time T to tumor onset is gamma distributed with the mean $\mu = 50$ years and the standard deviation $\sigma = 20$ years. The graph of the c.d.f. F_T is shown in Fig. 1. The law of tumor growth was taken to be deterministic exponential with the rate $\lambda = 1.6$ years^{-1}, which corresponds to the tumor size doubling time of approximately 5.2 months. The rate of spontaneous tumor detection was assumed to be $\alpha_0 = 0.03$. The graph of the survivor function $\bar{F}_{W_0}$ given by equation (9) is presented in Fig. 2. The effect of one exam occurring after tumor onset with the screening based tumor detection rate $\alpha = 0.1$ is shown in Fig. 3 featuring the survivor function of the time W to combined tumor detection.

The search for optimal screening schedules and optimal screening efficiencies was conducted for a fixed number n of screens with no restriction on the moments of exams and for various values of α. The method of optimization was the exhaustive search with the step 0.25 years. Parameter values $\mu = 50$ years, $\lambda = 1.6$ years^{-1}, and $\alpha_0 = 0.03$ were fixed throughout the calculations. For $n = 10$, plots of the rescaled optimal screening efficiency d with $\sigma = 20$ years versus α and, for $\alpha = 0.1$, versus σ are shown in Fig. 4 and 5, respectively. As it could be expected, d increases with increasing α and decreases with increasing σ.

The results of our search for optimal screening schedules with $n = 10,\ 20$ and with several values of σ and α are given in Table 1. For the reader's convenience, screening schedules are represented by the intervals $\Delta_i := \tau_i - \tau_{i-1},\ i = 1, ..., n$, between two successive exams. For all cases explored, optimal screening schedules are uniform or very close to such.

As a test for optimality of a screening schedule, profiles of the efficiency functional (22), with $n - 1$ moments of exams fixed at the optimal values and the remaining one varying between the two fixed neighboring moments of exams, were computed. For $n = 20$ and $\alpha = 0.1$, these profiles are given in Fig. 6. For the moment τ_1, a clear cut maximum was observed (see Fig. 6d), while for τ_{20} the maximum is more flat (see Fig. 6c). All intermediate moments of exams $\tau_2, ...\tau_{19}$ demonstrated a well-pronounced parabolic maximum (see Fig. 6a, b).

Acknowledgement

This work was supported by Grant DAMD17-98-1-8256 awarded by the U.S. Army Medical Research Acquisition Activity. The contents of this paper are solely the responsibility of the authors and do not necessarily reflect the position or the policy of the government, and no official endorsement should be inferred. The research of Dr. L.G. Hanin was also supported by Grant No. 807 from the Faculty Research Committee, Idaho State University, Pocatello, Idaho. We are very grateful to the reviewers whose comments have led to substantial improvements in the manuscript.

REFERENCES

1. A.Yu. Yakovlev and A.D. Tsodikov, *Stochastic Models of Tumor Latency and their Biostatistical Applications*, World Scientific Publ., Singapore (1996).

2. A.D. Tsodikov, B. Asselain, A. Fourquet, T. Hoang and A.Yu. Yakovlev, Discrete strategies of cancer post-treatment surveillance. Estimation and optimization problems, *Biometrics* **51**, 437-447 (1995).

3. A.Yu. Yakovlev, B. Asselain, V.-J. Bardou, A. Fourquet, T. Hoang, A. Rochefodiere and A.D. Tsodikov, A simple stochastic model of tumor recurrence and its application to data on premenopausal breast cancer. In *Biométrie et Analyse de Données Spatio-Temporelles* **12** (edited by B. Asselain, M. Boniface, C. Duby, C. Lopez, J.P. Masson, and J. Tranchefort), Société Française de Biométrie, ENSA, Rennes, 66-82 (1993).

4. R. Kirch and M. Klein, Surveillance schedules for medical examinations, *Management Science* **20**, 1403-1409 (1974).

5. A.K. Shahani and D.M. Crease, Towards models of screening for early detection of disease, *Adv. Appl. Prob.* **9**, 665-680 (1977).

6. D.M. Eddy, A mathematical model for timing repeated medical tests, *Medical Decision Making* **3**, 34-62 (1983).

7. A.D. Tsodikov and A.Yu. Yakovlev, On the optimal policies of cancer screening, *Math. Biosci.* **107**, 21-45 (1991).

8. A.D. Tsodikov, A.Yu. Yakovlev and L. Petukhov, Some approaches to screening optimization. In *Statistique des Processus en Milieu Médical* (edited by B. Bru, C. Huber and B. Prum) Université Paris V, Paris, France, 1-48 (1991).

9. A.D. Tsodikov, Screening under uncertainty. Games approach, *Syst. Anal. Model. Simul.* **9**, 259-262 (1992).

10. G. Parmigiani, On optimal screening schedules, *Biometric Bulletin* **8, 3**, 21 (1991).

11. G. Parmigiani, Optimal scheduling of fallible inspections, *DP 92-A38*, ISDS, Duke University (1992).

12. G. Parmigiani, On optimal screening ages, *J. Amer. Statist. Assoc.* **88**, 622-628 (1993).

13. G. Parmigiani, Timing medical examinations via intensity functions, *Biometrika* **84, 4**, 803-816 (1997).

14. G. Parmigiani and M.S. Kamlet, A cost-utility analysis of alternative strategies in screening for breast cancer. In *Bayesian Statistics in Science and Technology: Case Studies* (edited by C. Gatsonis, J. Hodges, R.E. Kass and N. Singpurwalla), Springer, New York, 390-402 (1993).

15. M. Zelen, Optimal scheduling of examinations for the early detection of disease, *Biometrika* **80**, 279-293 (1993).

16. Ping Hu and M. Zelen, Planning clinical trials to evaluate early detection programmes, *Biometrika* **84, 4**, 817-829 (1997).

17. S.J. Lee and M. Zelen, Scheduling periodic examinations for the early detection of disease: applications to breast cancer, *J. Amer. Statist. Ass.* **93**, 1271-1281 (1998).

18. L.V. Kantorovich and G.P. Akilov, *Functional Analysis*, 2nd ed., Pergamon Press, New York (1982).
19. S.S. Vallander, Calculation of the Wasserstein distance between probability distributions on the line, *Theory Prob. Appl.* **18**, 784-786 (1973).
20. S.T. Rachev and R.M. Short, Duality theorems for Kantorovich-Rubinstein and Wasserstein functionals, *Dissertationes Math.* **299** (1990).
21. L.G. Hanin, Kantorovich-Rubinstein norm and its application in the theory of Lipschitz spaces, *Proc. Amer. Math. Soc.* **115, 2**, 345-352 (1992).
22. S.H. Moolgavkar and D.J. Venzon, Two event model for carcinogenesis: Incidence curves for childhood and adult tumors, *Math. Biosci.* **47**, 55-77 (1979).
23. S.H. Moolgavkar and A.G. Knudson, Mutation and cancer: a model for human carcinogenesis, *J. Natl. Cancer Inst.* **66**, 1037-1052 (1981).
24. W. F. Heidenreich, On the parameters of the clonal expansion model, *Radiat. Envir. Biophys.* **35**, 127-129 (1996).
25. L.G. Hanin and A.Yu. Yakovlev, A nonidentifiability aspect of the two-stage model of carcinogenesis, *Risk Analysis* **16, 5**, 711-715 (1996).
26. W.F. Heidenreich, E.G. Luebeck and S.H. Moolgavkar, Some properties of the hazard function of the two-mutation clonal expansion model, *Risk Analysis* **17**, 391-399 (1997).
27. A. Kopp-Schneider, C.J. Portier and C.D. Sherman, The exact formula for tumor incidence in the two-stage model, *Risk Analysis* **14**, 1079-1080 (1994).
28. Q. Zheng, On the exact hazard and survival functions of the MVK stochastic carcinogenesis model, *Risk Analysis* **14**, 1081-1084 (1994).
29. A.Yu. Yakovlev and E. Polig, A diversity of responses displayed by a stochastic model of carcinogenesis allowing for cell death, *Math. Biosci.* **132**, 1-33 (1996).
30. A.Yu. Yakovlev, L.G. Hanin, S.T.. Rachev and A.D. Tsodikov, Distribution of tumor size at detection and its limiting form, *Proc. Natl. Acad. Sci. USA* **93**, 6671-6675 (1996).
31. L.G. Hanin, S.T. Rachev, A.D. Tsodikov and A.Yu. Yakovlev, A stochastic model of carcinogenesis and tumor size at detection, *Adv. Appl. Prob.* **29**, 607-628 (1997).
32. L.G. Hanin and K.M. Boucher, Identifiability of parameters in the Yakovlev-Polig model of carcinogenesis, *Math. Biosci.* (1999), to appear.
33. R. Bartoszyński, A modeling approach to metastatic progression of cancer. In *Cancer Modeling* (edited by J.R. Thompson and B.W. Brown), Marcel Dekker, New York and Basel, 237-267 (1987).
34. M. Klein and R. Bartoszyński, Estimation of growth and metastatic rates of primary breast cancer. In: *Mathematical Population Dynamics* (edited by O. Arino, D. E. Axelrod and M. Kimmel), Marcel Dekker, New York, 397-412 (1991).
35. N.E. Atkinson, R. Bartoszyński, B.W. Brown and J.R. Thompson, On estimating the growth function of tumors, *Math. Biosci.* **67**, 145-166 (1983).
36. N.E. Atkinson, B.W. Brown and J.R. Thompson, On the lack of concordance between primary and secondary tumor growth rates, *J. Natl. Cancer Inst.* **78**, 425-435 (1987).
37. B.W. Brown, N.E. Atkinson, R. Bartoszyński and E.D. Montague, Estimation

of human tumor growth rate from distribution of tumor size at detection, *J. Natl. Cancer Inst.* **72**, 31-38 (1984).

Table 1. **Optimal screening schedules**

n=20

σ	α	$\Delta_1=\tau_1$	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7	Δ_8	Δ_9	Δ_{10}
20	0.1	27.75	2.25	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.25
		Δ_{11}	Δ_{12}	Δ_{13}	Δ_{14}	Δ_{15}	Δ_{16}	Δ_{17}	Δ_{18}	Δ_{19}	Δ_{20}
		2.25	2.25	2.25	2.25	2.25	2.25	2.25	2.25	2.25	2.25

n=10

σ	α	$\Delta_1=\tau_1$	Δ_2	Δ_3	Δ_4	Δ_5	Δ_6	Δ_7	Δ_8	Δ_9	Δ_{10}
20	0.1	35.00	2.50	2.50	2.50	2.25	2.50	2.50	2.50	2.50	2.50
20	0.3	35.00	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50	2.50
20	0.5	34.50	2.75	2.75	2.50	2.50	2.50	2.50	2.50	2.75	2.75
10	0.1	40.75	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.00	2.25
30	0.1	27.25	2.50	2.75	2.50	2.50	2.50	2.50	2.50	2.50	2.50

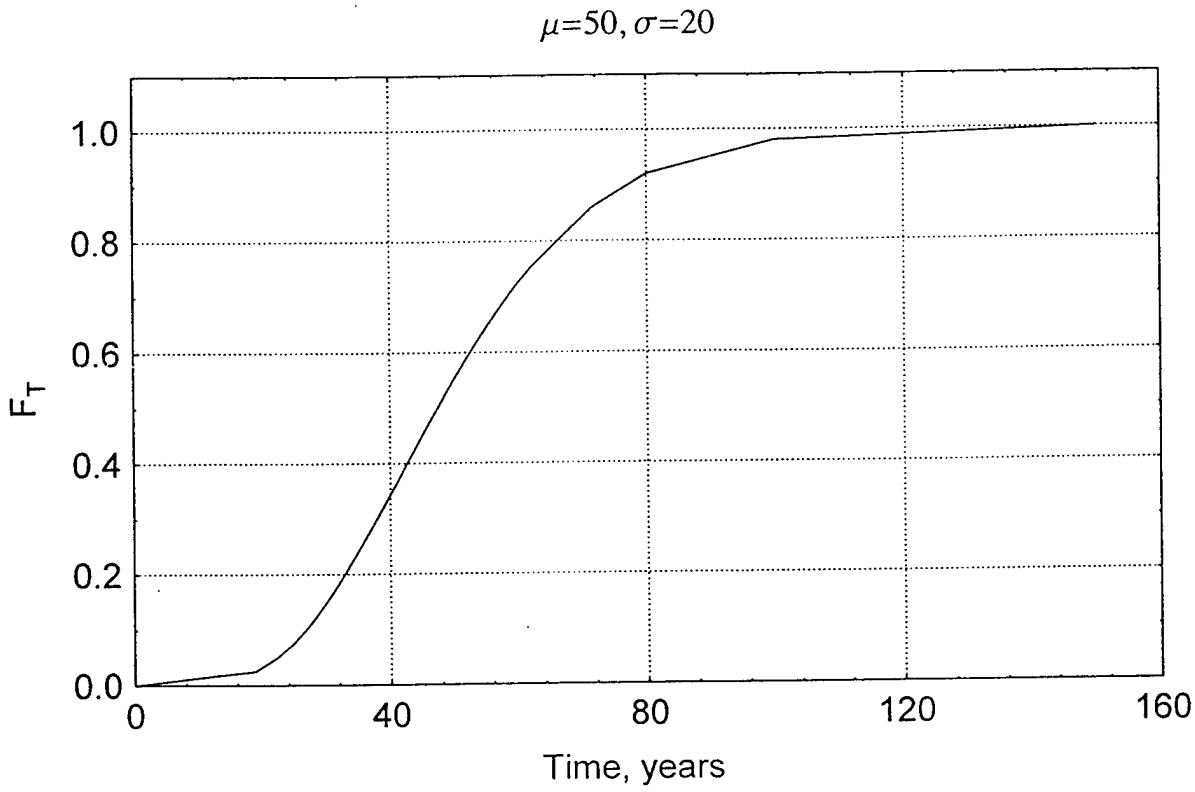

Figure 1. The cumulative distribution function for the time to tumor onset.

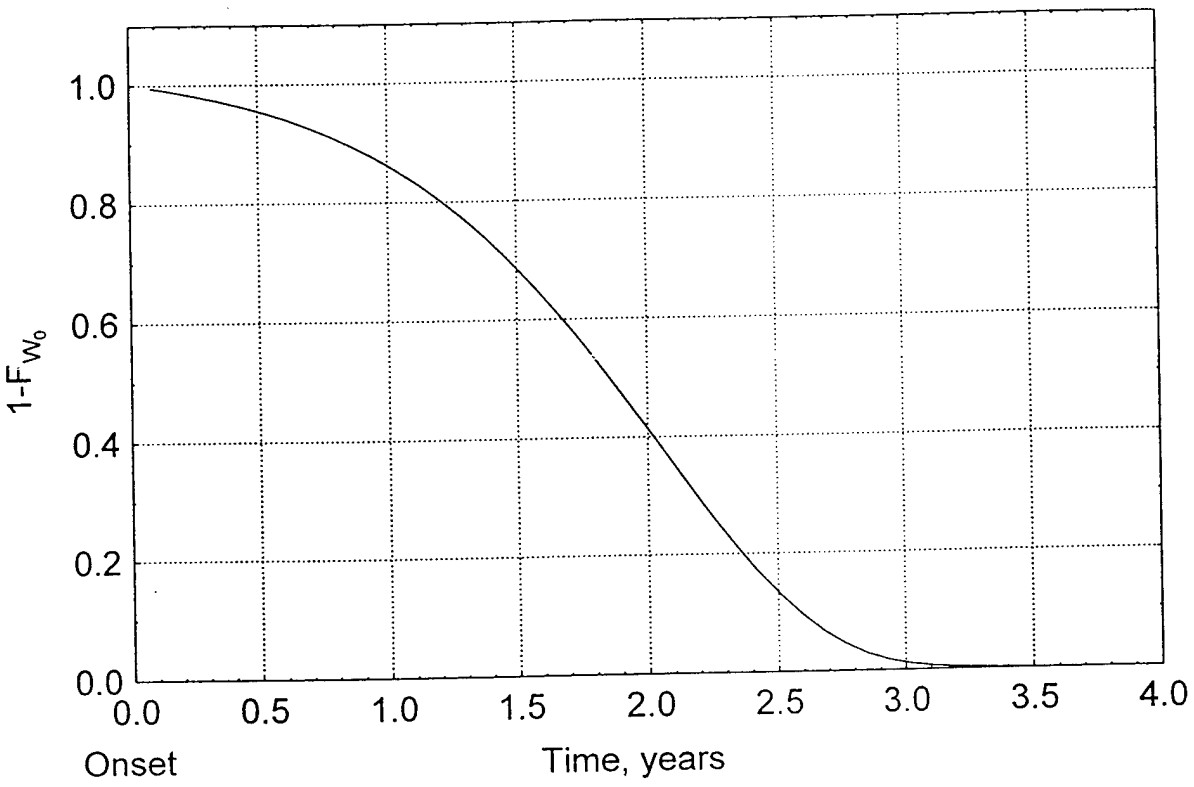

Figure 2. The survivor function for the time to spontaneous detection.

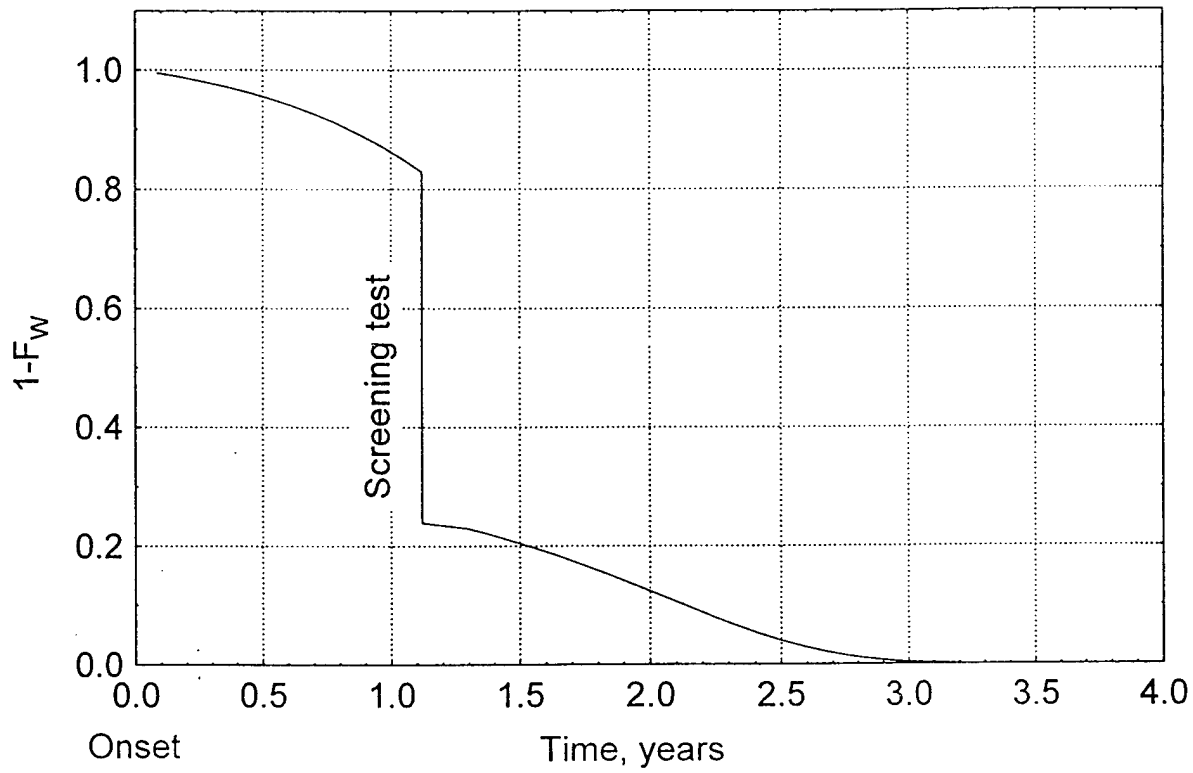

Figure 3. The survivor function for the time to combined detection ($\alpha = 0.1$).

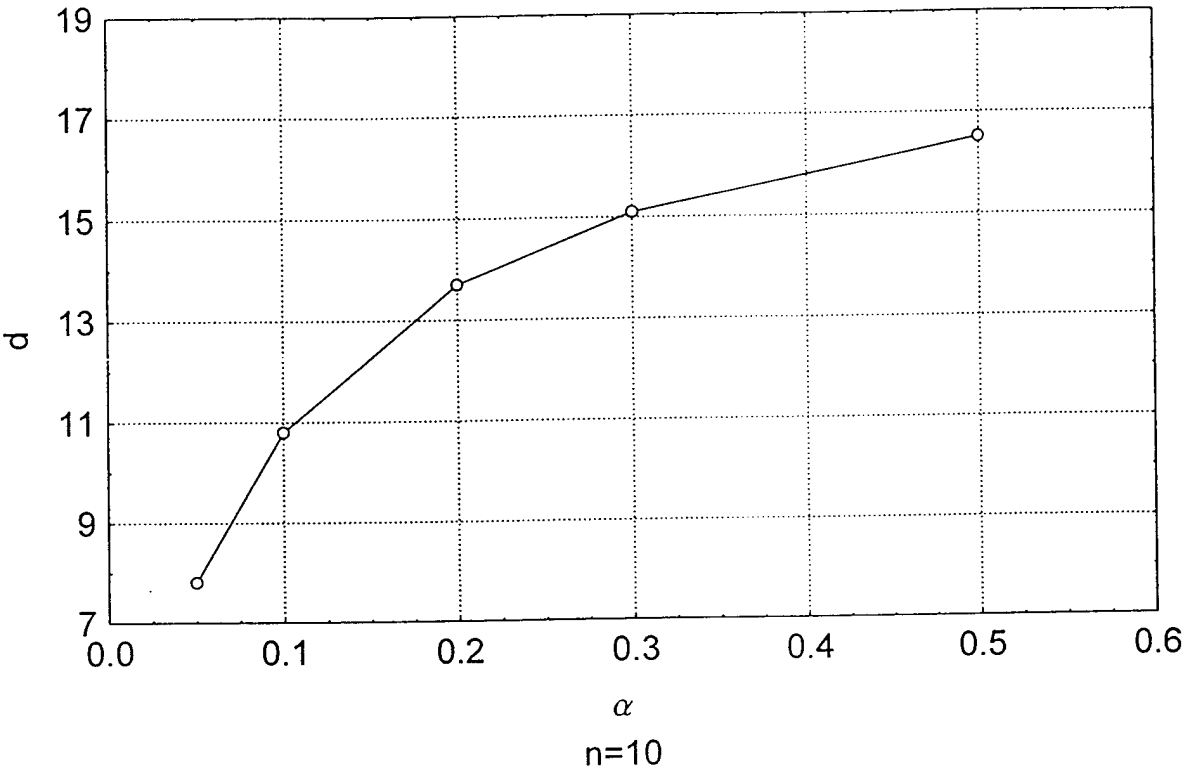

Figure 4. Optimal screening efficiency as a function of the parameter α given a fixed number ($n = 10$) of examinations and $\sigma = 20$ years.

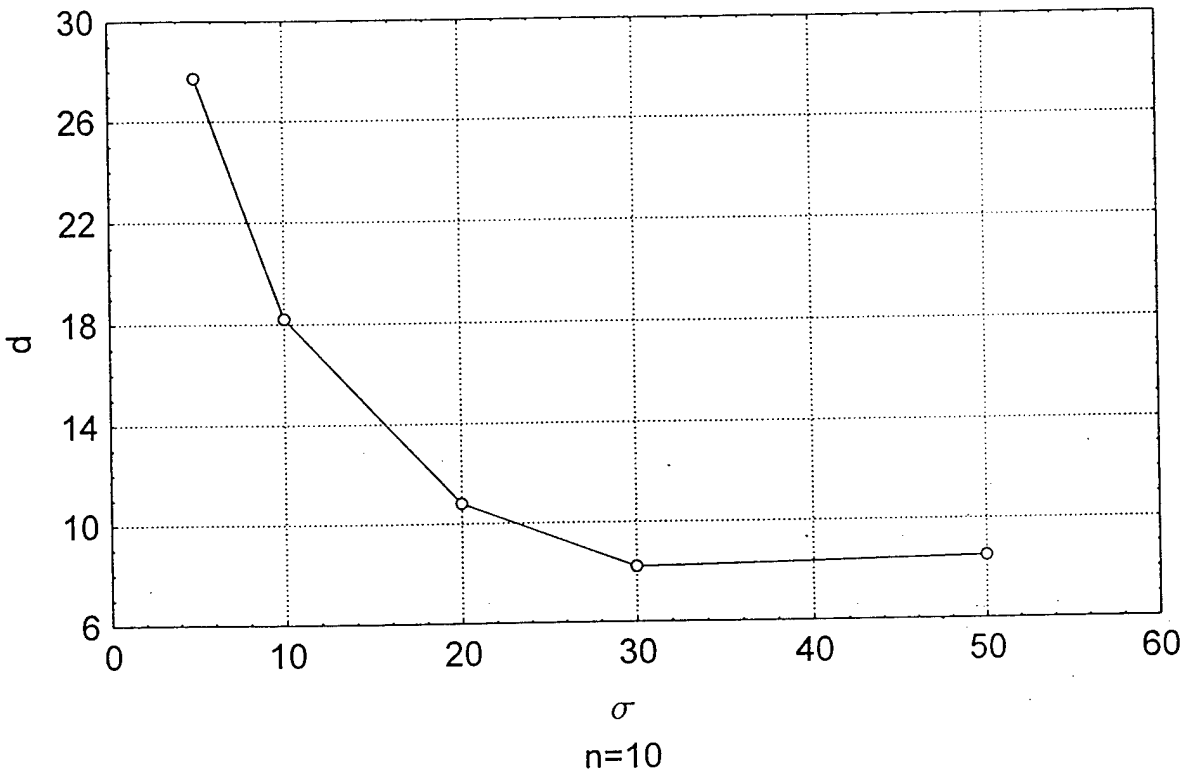

Figure 5. Optimal screening efficiency as a function of the standard deviation, σ, of the onset time ($n = 10$, $\alpha = 0.1$).

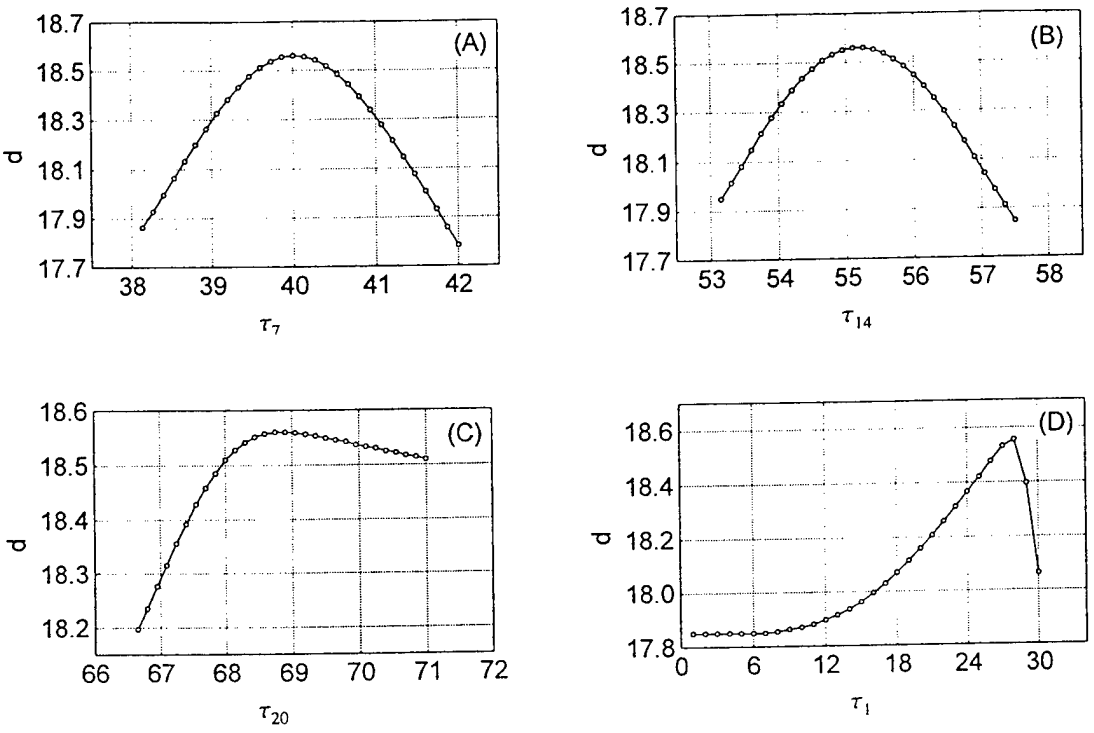

Figure 6. Profiles of the efficiency functional ($n = 20, \sigma = 20$ years, $\alpha = 0.1$).