Chapter 10:

The University of Rochester Model of Breast Cancer Detection and Survival

Leonid G. Hanin, Anthony Miller, A. V. Zorin, Andrei Y. Yakovlev

This paper presents a biologically motivated model of breast cancer development and detection allowing for arbitrary screening schedules and the effects of clinical covariates recorded at the time of diagnosis on posttreatment survival. Biologically meaningful parameters of the model are estimated by the method of maximum likelihood from the data on age and tumor size at detection that resulted from two randomized trials known as the Canadian National Breast Screening Studies. When properly calibrated, the model provides a good description of the U.S. national trends in breast cancer incidence and mortality. The model was validated by predicting some quantitative characteristics obtained from the Surveillance, Epidemiology, and End Results data. In particular, the model provides an excellent prediction of the size-specific age-adjusted incidence of invasive breast cancer as a function of calendar time for 1975–1999. Predictive properties of the model are also illustrated with an application to the dynamics of age-specific incidence and stage-specific age-adjusted incidence over 1975–1999. [J Natl Cancer Inst Monogr 2006;36:66–78]

The purpose of the modeling effort described herein is twofold:

- Designing explanatory and predictive tools for quantitative description of the effects of breast cancer screening for various screening strategies, including the national trends in breast cancer incidence and mortality under the base case scenario
- Developing methods for statistical inference on the natural history of breast cancer in terms of biologically meaningful parameters

In what follows, we use the term "prediction" to mean extrapolation of the basic epidemiological descriptors from one setting to another, including new interventions and risk factors, but not the problem of forecasting future population trends. The latter sort of model-based predictions would require sufficient knowledge of future changes in all components of the natural history of the disease, including future changes of cancer risk over time. See *(1)* for the discussion of this problem in regard to the age–period–cohort model.

The traditional approach to mathematical or simulation modeling of cancer screening tends to describe the process of tumor development in only one dimension, that is, the *time natural history*. A broader methodological idea is to construct a stochastic model of cancer development and detection that yields the multivariate distribution of observable variables at the time of diagnosis *(2)*. By focusing on such multivariate observations, rather than just on the age of patients at diagnosis, this idea seeks to invoke an additional source of information (available only at the time of detection) to

improve estimation of unobservable parameters of cancer latency. Indeed, the process of tumor progression manifests itself as certain changes in many characteristics of the tumor. Therefore, this process is multidimensional in nature, and modeling tumor progression as a (linear) sequence of stages represents a poor approximation to a more general multivariate model of the natural history of cancer. The idea of multiple pathways of cancer progression was introduced in the path-breaking works by *(3)* and Feldstein and Zelen *(4)*. In this paper, we base our inference on the natural history of breast cancer on two important variables, namely, the tumor size and the age of a patient at diagnosis, using mechanistic models of tumor development and detection to derive an analytic expression of the joint distribution of the said variables. In doing so, we take advantage of a mechanistic two-stage model of carcinogenesis to describe the "disease-free" stage of breast cancer development and the so-called quantal response model to relate a chance of detecting a tumor to its size; the latter mechanism applies equally to both incident and screen-detected (prevalent) cases. Some authors *(5–7)* have long realized the advantages of multivariate analysis in screening studies, but specific modeling and inferential techniques require a much higher level of sophistication than that in the earlier attempts at a comprehensive theory of cancer screening.

The proposed model of the natural history of breast cancer has the following advantages:

1. It is based on a minimal set of biologically plausible assumptions.
2. It is proven to be completely identifiable.
3. It is formulated in terms of probabilistic characteristics that can be estimated in the presence of data censoring, thereby requiring no demographic information for their evaluation.
4. When applied to the data generated by randomized screening trials, the model allows estimation of all parameters by the method of maximum likelihood, whereas a subset of its parameters responsible for the progression (preclinical) stage of tumor development can independently be estimated from the population-based data available from the Surveillance, Epidemiology and End Results (SEER) National Cancer Institute (NCI) program.
5. All parameters are estimated from epidemiological data, using the same model that makes them by far more reliable

Affiliations of authors: Department of Mathematics, Idaho State University, Pocatello, ID (LGH); Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY (LGH, AVZ, AYY); Department of Public Health Sciences, University of Toronto, Toronto, Ontario, Canada (AM).

Correspondence to: Andrei Y. Yakovlev, PhD, Department of Biosciences, University of Rochester, 601 Elamwood Ave., Box 630, Rochester, NY 14642 (e-mail: andrei_yakovlev@urmc.rochester.edu).

See "Notes" following "References."

than those available from the literature, because the latter estimates have been obtained under dramatically dissimilar models and assumptions.

6. When extrapolations are made to a different dataset, minimal calibration is needed, involving only those parameters that are likely (on biological grounds) to vary between the two sets of data.

7. The predictive power of the model has been evaluated in several applications under strict conditions allowing no further calibration of any of its parameters already estimated (calibrated) in a different setting.

The model has been built in part on the base case inputs shown in Table 1.

The natural history model allows us to estimate the effects of screening on the age-specific cancer incidence and the distribution of major covariates at the time of diagnosis. This inference is entirely *independent of the data on cancer mortality*. To model the effect of screening on cancer-specific mortality, one needs to establish a quantitative relationship between clinical covariates (e.g., age, stage, tumor size) and postdetection survival of patients with breast cancer. Regression survival models are designed to estimate the survival time distribution conditional on covariate information, whereas the joint (multivariate) distribution of covariates at the time of diagnosis provides a link between the natural history of breast cancer and cancer-specific survival. The periodic screening evaluation methodology *(8)*, although elegant, does not represent a strong alternative to flexible natural history models because of its many limitations—among which is the assumption that any effect of birth cohort is negligible. Our analysis of the Utah Population Database have shown that breast cancer risk varies substantially between birth cohorts separated by time intervals as short as 5 years.

We resorted to a new class of extended hazard regression models with cure that has been extensively studied in recent years *(9–15)*, to name a few. Even the simplest model from this class provides an excellent description of the effects of clinical covariates on cancer survival *(13,16,17)*. In combination with the natural history component, this model was used in our studies to

Table 1. Base case parameter usage*

Parameter	Usage
Base case treatment dissemination	Not needed
Base case mammography dissemination	Used in provided form
Base case other-cause mortality	Not needed
Base case age-specific breast cancer incidence	Used for validation
Base case age-adjusted breast cancer incidence	Some values were used for calibration
Base case 1975 breast cancer prevalence	Not needed
Base case 1975 cause specific survival	Not needed
Base case historical survival	Not needed
Base case 1975 breast cancer mortality	Used for calibration
Base case breast cancer APC incidence	Uses a processed version of the standard parameter (relative risks)
Base case treatment effect	Not needed
Base case SEER 9 mortality	Used for calibration

*APC = age–period–cohort; SEER = Surveillance, Epidemiology, and End Results.

describe the U.S. national trend in breast cancer mortality from 1975 to 1999. The observed mortality trend is consistent with the assumption that there has been a relatively long history of incremental improvements in breast cancer treatment. The conjecture is corroborated by a recent analysis of breast cancer mortality in the United Kingdom *(18)* indicating that the mortality rate began to decrease before the start of mammographic screening.

The model has proven to be adequate for the complex phenomena that so far have been explored in relation to cancer incidence and mortality. This statement, however, does not mean that our model is either perfect or universal; it may call for further modifications if future applications so require. We believe that improved models of cancer screening can be developed in the future through including more components (in addition to age and tumor size) in the vector of clinical covariates accessible to measurement at the time of diagnosis. See *(2)* for the general idea and associated analytical techniques.

MODELING THE NATURAL HISTORY OF BREAST CANCER

Our approach attempts to implement the following concept formulated by Albert et al. *(19)*: More realistic models for tumor detectability can be synthesized by first modeling the behavior of tumor growth over time and superimposing a model for detection probability as a function of tumor size. This idea was set forth in the so-called quantal response model of tumor detection developed by Bartoszyński and other authors in several publications [see *(20)* for details and references]. Below we outline the most basic features of the class of quantal response models of cancer detection combined with mechanistically motivated models of carcinogenesis.

A Stochastic Model of Tumor Latency

The latent period of tumor development can be broken down into three stages:

1. formation of initiated cells,
2. promotion of initiated cells resulting in the first malignant clonogenic cell,
3. subsequent tumor growth and progression until the event of detection occurs.

Let T be the age of an individual at tumor onset, and W the time of spontaneous tumor detection (in the absence of screening) measured from the onset of disease. We use a two-stage stochastic model of carcinogenesis proposed by Moolgavkar et al. *(21–23)* to specify the probability density function, $p_T(t)$, of the random variable T. The model is given by the following survival function

$$\overline{F}_T(t) := \int_t^\infty p_T(u)du = \left[\frac{(A+B)e^{At}}{B+Ae^{(A+B)t}}\right]^\rho, \quad t \geq 0, \ [10.1]$$

from which the density $p_T(t)$ can be derived. Here $A, B, \rho > 0$ are identifiable parameters of the model *(24–26)*. Formula [10.1] specifies the distribution of the duration of the first two stages of carcinogenesis. The process of initiation is usually modeled as a Poisson process. Here the parameter ρ is the ratio of the initiation rate and the rate of proliferation of initiated cells, whereas A and B are parameters of the promotion time distribution. This model has proven to provide a good fit to diverse data on animal and human carcinogenesis [see *(27)* for goodness-of-fit testing].

Introduce a random variable S to represent tumor size (the number of cells in a tumor) at spontaneous detection. Suppose that the

law of tumor growth is described by a deterministic function $f:[0,\infty)\to[1,\infty)$ with $f(0)=1$, so that $S=f(W)$. It is assumed also that

1. the random variables T and W are absolutely continuous and stochastically independent;
2. the function f is differentiable and $f'>0$; and
3. the hazard rate for spontaneous tumor detection is proportional to the current tumor size with coefficient $\alpha>0$.

It follows from the above assumptions that

$$p_W(w) = \alpha f(w)e^{-\alpha \int_0^w f(u)du}, \quad w \geq 0.$$

Therefore,

$$p_S(s) = \alpha s g'(s)e^{-\alpha \int_0^{g(s)} f(u)du}, \quad s \geq 1,$$

where g stands for the inverse function for f: $g = f^{-1}$. For deterministic exponential tumor growth with rate $\lambda > 0$ ($f(w) = e^{\lambda w}$), we have

$$p_S(s) = \frac{\alpha}{\lambda}e^{-\frac{\alpha}{\lambda}(s-1)}, \quad p_W(w) = \alpha e^{\lambda w - \frac{\alpha}{\lambda}(e^{\lambda w}-1)},$$
$$s \geq 1, \; w \geq 0. \qquad [10.2]$$

Here tumor size at detection S follows a translated exponential distribution with parameter α/λ, whereas the distribution of age at detection measured from the disease onset is a Gompertz distribution. The random variable W has the same meaning as the pre-clinical stage duration within the traditional approach to modeling the natural history of cancer.

Consider the random vector $Y: = (T + W, S)$, which components are interpreted as age and tumor size at diagnosis, respectively. The probability density function of Y is given by

$$p_Y(u,s) = p_T(u - g(s))p_S(s), \quad u \geq g(s), s \geq 1.$$

This distribution is identifiable *(28)*.

Remark 1. A distribution $P(x;\theta)$, where θ is a vector of parameters, is said to be identifiable if from the equality

$$P(x; \theta_1) = P(x; \theta_2)$$

valid for all x it follows that $\theta_1 = \theta_2$.

For exponential tumor growth, we obtain

$$p_Y(u,s) = \frac{\alpha}{\lambda}e^{-\frac{\alpha}{\lambda}(s-1)}p_T(u - \frac{\ln s}{\lambda}), \quad u \geq 0, \; 1 \leq s \leq e^{\lambda u}.$$

In practice, it is not the number of tumor cells S that is observable but the volume, V, in appropriate units, and one needs to change variables using the equality $S = V/\gamma$, where γ is the volume of one tumor cell ($\gamma \approx 10^{-9}\text{cm}^3$). The parameter γ has only a scaling effect on the distribution of tumor volume.

The most straightforward generalization of the model can be accomplished by assuming that some of its parameters are random. In particular, suppose that $1/\lambda$ is gamma distributed with parameters a and b. Randomness of λ is reflective of the individual variability of tumor growth rate. Then we have

$$p_Y(u,s) = \frac{\alpha b^a}{(\ln s)^{a+1}\Gamma(a)} \int_0^u (u-x)^a$$
$$\times \exp\left\{-\frac{b+\alpha(s-1)}{\ln s}(u-x)\right\}p_T(x)dx, \qquad [10.3]$$

for $u \geq 0$, $s \geq 1$. Here the marginal distribution of tumor size is a Pareto distribution *(20)*. As shown by our analysis of the Utah

Population Data Set, the distribution [10.3] provides a good fit to cohort data on breast cancer *(29)*.

Modeling Impact of Screening on Natural History of Breast Cancer

Let $0 < \tau_1 < \tau_2 < \ldots < \tau_n$ be a given screening schedule. It is convenient to set $\tau_0: = 0$ and $\tau_{n+1}: = \infty$. Let W_0 be the time of spontaneous detection (incident and interval cases) and W_1 the time of screening-based detection, both times being measured from the disease onset. Then for the time W of combined detection we have $W = \min(W_0, W_1)$. It is natural to assume that *the random variables W_1 and W_0 are conditionally independent given the onset time T*. This assumption is plausible for deterministic tumor growth if we hypothesize also that W_0 and T are independent and that the hazard rate for the distribution of W_0 as well as the discrete hazard rate of tumor detection at a medical exam (provided the previous exams did not detect the tumor) given T are both proportional (with different coefficients of proportionality α_0 and α_1, respectively) to the current tumor size. Then we have for $u \geq 0, s \geq 1$:

$$F_Y(u,s) = \Pr(T + W \leq u, f(W) \leq s)$$
$$= \int_0^u F_{W \mid T=t}(\min\{u - t, g(s)\})f_T(t)dt, \quad [10.4]$$

where

$$F_{W \mid T=t}(w) = 1 - e^{-[\alpha_0\Phi(w)+\alpha_1 \sum_{k=i+1}^{j} f(\tau_k-t)]}.$$

It can be proven that under mild conditions the distribution of the random vector Y is identifiable *(30)*. Unfortunately, no identifiability results are available for its randomized versions because of prohibitive complexity of the analytic expression for this distribution even for exponential tumor growth with random growth rate. The same is true for the joint distribution given by formula [10.3]. However, our simulation experiments have shown that identifiability of the model is preserved if the compounding procedure uses the gamma distribution for λ or its reciprocal.

Remark 2. It is tempting to generalize the model by making the growth rate (or the preclinical stage duration) dependent on the age of a patient at the time of tumor onset. There have been attempts to incorporate this element into models of the natural history of cancer. However, the fact that no tangible birth cohort effect on the nonparametrically estimated distribution of tumor size at detection is seen in a cohort study *(29)* is consistent with stochastic independence of tumor growth rate of the age at tumor onset. The same argument applies equally to the sensitivity parameter α_1. There are indications in the literature *(31)* that the sensitivity of mammography increases with age. However, the tendency does not appear to be strong enough to manifest itself in the type of data we deal with in this paper.

Next we need a formula for the probability of detection at a given screen. 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 kth 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. Introduce a discrete analogue of the conditional (given $T = t$) hazard rate for the screening based detection as

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

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. Then the following formula holds *(32)*:

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

Observe that this holds true for all $k = 1, \ldots, n$, if we set $pt(k) = r_t(k) = 0$ for $1 \leq k \leq i$. Assuming also that the conditional discrete rate of screening based detection is proportional to the current tumor size: $r_t(k) = \alpha_1 S(\tau_k - t)$, $i+1 \leq k \leq n$, $\alpha_1 > 0$, we have

$$p_t(k) = e^{-\sum_{j=i+1}^{k-1} \alpha_1 S(\tau_j - t)}[1 - e^{-\alpha_1 S(\tau_k - t)}], \quad i+1 \leq k \leq n. \quad [10.5]$$

Estimation of Model Parameters

When considering a study design typical for randomized screening trials it is possible to derive the likelihood function on the basis of the observations of tumor size and age at diagnosis. Let $\tau = \{\tau_1 < \tau_2 < \ldots < \tau_n\}$ be a sequence of screening ages. We proceed from the model assumptions introduced earlier with an arbitrary probability density function $p_\Lambda(x)$ for the rate, λ, of tumor growth. The model is formulated in terms of tumor size and age at detection denoted by S and U, respectively. The contributions to the likelihood function are given by the following formulas *(30)*:

1. *For interval and incident cases,* the contribution of an observation (u, s) is equal to the value at (u, s) of the joint p.d.f. of age and tumor size at spontaneous (clinical) detection:

$$p_1(u, s) = \alpha_0 \int_0^{u-\tau_k} \exp\{-\frac{\alpha_0 x}{\ln s}(s-1)]\}p_T(u-x)p_\wedge(\frac{\ln s}{x})$$

$$\times \frac{dx}{x} + \alpha_0 \sum_{i=0}^{k-1} \int_{u-\tau_{i+1}}^{u-\tau_i} \exp\{-[\frac{\alpha_0 x}{\ln s}(s-1) + \alpha_1 s$$

$$\times \sum_{j=i+1}^{k} e^{\frac{\ln s}{x}(\tau_j - u)}]\}p_T(u-x)p_\wedge(\frac{\ln s}{x})\frac{dx}{x}, \quad [10.6]$$

where $\tau_k \leq u < \tau_{k+1}$, $0 \leq k \leq n$.

2. *For screen-detected cases,* the contribution of an observation (τ_k, s), $1 \leq k \leq n$, equals the value of the p.d.f. of tumor size at detection on the kth screen:

$$p_2(\tau_k, s) = \frac{1 - e^{-\alpha_1 s}}{s} \sum_{i=0}^{k-1} \int_{\tau_k - \tau_{i+1}}^{\tau_k - \tau_i} \exp\{-\frac{\alpha_0(s-1)x}{\ln s}\}$$

$$\times \exp\{-\alpha_1 s \sum_{j=i+1}^{k-1} e^{-\frac{\ln s}{x}(\tau_k - \tau_j)}\}p_T(\tau_k - x)$$

$$\times p_\wedge(\frac{\ln s}{x})\frac{dx}{x}, \quad 1 \leq k \leq n. \quad [10.7]$$

3. *For censored observations,* the contribution of an observation c equals $\overline{F}_U(c)$, where U is the age at combined tumor detection:

$$p_3(c) = \overline{F}_T(c) + \int_0^\infty \int_0^{c-\tau_k} \exp\{-\frac{\alpha_0}{\lambda}(e^{\lambda x} - 1)\}$$

$$\times p_T(c-x)p_\wedge(\lambda)dx d\lambda$$

$$+ \sum_{i=0}^{k-1} \int_0^\infty \int_{c-\tau_{i+1}}^{c-\tau_i} \exp\{-[\frac{\alpha_0}{\lambda}(e^{\lambda x} - 1)$$

$$+ \alpha_1 e^{-\lambda(c-x)} \sum_{j=i+1}^{k} e^{\lambda \tau_j}]\}p_T(c-x)p_\wedge(\lambda)dx d\lambda, \quad [10.8]$$

Here $\overline{F}_T$ is the survival function of the onset time, and $\tau_k \leq c < \tau_{k+1}$, $0 \leq k \leq n$. Since age at enrollment varies among study sub-

jects, the above formulas must be modified in the usual way in order to incorporate left random truncation *(33)*.

We estimated all parameters incorporated into the model from individual data generated by the Canadian National Breast Screening Studies (CNBSS). This study consists of two screening trials individually randomized and conducted during 1980–96 and monitored to 1996 in 15 centers in Canada. Both trials were coordinated at the University of Toronto by a study team directed by one of the coauthors of this paper (Miller).

The first study (CNBSS-1) included 50 430 women aged 40–49 years on study entry and evaluated the efficacy of annual mammography, breast physical examination, and breast self-examination instruction (BSE) in reducing breast cancer mortality *(34)*. In the mammography plus physical examination group, 62% of women received five annual screens, including two-view mammography, physical examination, and BSE. The remaining women, recruited later, received four annual screens. The usual care group were not recalled for rescreening after their first visit, when they had a breast physical examination plus BSE, but were mailed annual questionnaires. The second study (CNBSS-2) included 39 405 women age 50–59 years on study entry and evaluated the contribution of annual mammography over and above annual physical examination of the breasts plus BSE in the reduction of mortality from breast cancer *(34)*. Women were randomized to receive annual mammography and physical examination plus BSE or annual physical examination plus BSE only for a total of five or four screens.

For both trials, center coordinators conducted the randomization using allocation lists prepared by the central office. Randomization was independent of physical examination findings. The center coordinators collected surgery and pathology reports for all breast diagnostic and therapeutic procedures. CNBSS pathologists reviewed all slides. If the community and CNBSS pathologist disagreed, a panel of three to five CNBSS pathologists conducted a blind and independent review. Extensive quality-control procedures were carried out while data collection was in progress. After the screening centers closed in 1988, all women known to have breast cancer were followed up annually by the CNBSS central office until June 30, 1996. Passive follow-up of all participants through linkage with the National Cancer Registry identified new diagnoses of breast cancer in study participants to December 31, 1993. The central office collected pathology reports for postscreen breast cancers. For these, the community diagnosis was accepted for study purposes. Deaths that occurred before a participant's screening schedule was completed were identified by family members in response to the annual mailed questionnaire. Attending physicians, who received annual requests for information on women with breast cancer, reported deaths to June 30, 1996. Linkage with the CMDB (including deaths in Canadians resident in the United States at time of death) identified causes of death in the entire cohort to December 31, 1993. Independent reviewers, blind as to allocation, reviewed clinical records and classified the underlying cause of death.

In CNBSS-1, a total of 592 invasive and 71 in situ breast cancers were diagnosed by December 31, 1993, in the mammography plus physical examination group, compared with 552 and 29, respectively, in the usual care group. Of these, 208 and 58, respectively, were screen detected. At 7 years, there were 38 breast cancer deaths in the mammography plus physical examination group and 28 in the usual care group, for a rate ratio of

1.36 *(35)*. At the 11- to 16 (average 13)-year follow-up, there were 105 and 108 breast cancer deaths, respectively, for a rate ratio, adjusted for mammograms performed outside the CNBSS, of 1.06 *(36)*. In CNBSS-2, a total of 622 invasive and 71 in situ breast carcinomas were ascertained in the mammography plus physical examination group, and 610 and 16 in the physical examination only group. Of these, 267 and 148, respectively, were screen detected. At 7 years there were 38 and 39 deaths from breast cancer in the respective groups for a rate ratio of 0.97 *(37)*. At 11–16 years there were 107 and 105 deaths from breast cancer in the respective groups for a rate ratio of 1.02 *(38)*. Information on tumor size and clinical stage at diagnosis is available in both datasets.

Maximization of the likelihood given by [10.6]–[10.8] is a challenging problem because it involves many time-consuming computations. This statement is especially true for many (tens of thousands) double integrals [10.8] representing the contributions of censored observations, for censoring is heavy in this kind of study. Therefore, we resorted to simulations to estimate the contributions of censored data rather than evaluate these integrals numerically. The simulation model described in the next section was used for this purpose. The survival function $\bar{F}_U$ (for censored observations) and the p.d.f. p_U (for missing tumor size) were estimated nonparametrically from the simulated data. There is always a certain level of random noise in the simulated likelihood, calling for stochastic approximation methods to find a maximum of its expected value. Therefore, we used the Kiefer–Wolfowitz procedure *(39)* to obtain maximum likelihood estimates. Unfortunately, when applied to the log-likelihood function the Kiefer–Wolfowitz procedure may result in biased estimates and one needs to generate extremely large simulation samples to keep this bias to a minimum. For this reason, we provided 10^5 simulated samples when estimating $\bar{F}_U(u_i)$ and 5×10^5 samples when estimating $p_U(u_i)$ per each iteration of the Kiefer–Wolfowitz procedure in the likelihood inference from the Canadian screening trials. In a separate set of simulation experiments, we assured ourselves that this sample size was sufficient for obtaining stable results. The contributions of exact observations were computed numerically in accordance with formulas [10.6] or [10.7].

In our analysis of the CNBSS data, we assumed that the reciprocal of Λ in formulas [10.6]–[10.8] is gamma distributed with mean μ and standard deviation σ. Since there are three modes of breast cancer detection in the trial, we extended the likelihood function to incorporate three sensitivity parameters: α_0, α_1, and α_2 for spontaneous (clinical) detection, mammography combined with physical exam, and physical exam only, respectively. From age at enrollment one can pick out four distinct birth cohorts in the CNBSS data. Each of the four cohorts is composed of women born within the following 5-year intervals: 1921–25, 1926–30, 1931–35, and 1936–40, respectively. If we allow the parameter ρ to vary among the different birth cohorts, there will be four parameters ρ_1, ρ_2, ρ_3, ρ_4 forming the proportional hazards structure of the onset time distribution. We made these parameters responsible for the birth cohort effect as suggested by Boucher and Kerber *(40)*. The remaining two parameters A and B (see formula [10.1]), along with the parameters μ, σ, α_0, α_1, and α_2, are assumed to be common to all birth cohorts. Therefore, there are 11 parameters in total to be estimated from the CNBSS data by the method of maximum likelihood.

Our procedure resulted in the following maximum likelihood estimates (MLE) of ρ_1, ρ_2, ρ_3, ρ_4: $\hat{\rho}_1 = 0.059$, $\hat{\rho}_2 = 0.063$, $\hat{\rho}_3 = 0.084$,

$\hat{\rho}_4 = 0.09$. These estimates indicate that breast cancer risk tends to increase over the time range covered by the birth cohorts under study. The MLEs of the parameters A, B, μ, σ and their 95% confidence intervals are given in Table 2. The construction of approximate confidence intervals is based on asymptotic normality of maximum likelihood estimators. The MLEs of α_0, α_1, and α_2 are equal to 7.31×10^{-10}, 4.82×10^{-9}, and 1.34×10^{-9}, respectively, with the corresponding confidence intervals: (6.91×10^{-10} to 7.72×10^{-10}), (4.45×10^{-9} to 5.18×10^{-9}), and (1.24×10^{-9} to 1.44×10^{-9}). There is a more than threefold difference in the sensitivity parameters associated with mammography combined with physical exam (α_1) and with physical exam alone (α_2), respectively.

Randomized screening trials are especially well suited for estimation of model parameters in a statistically sound way, because such studies generate individual data on age and tumor size at detection. However, it is the objective of our study to provide a means of explanatory and predictive inference at the population level. We use the CNBSS data to estimate parameters associated with the four birth cohorts identified through the Canadian studies, keeping in mind that some of these parameters may still be adjusted when an additional calibration of the model is necessary (see "Model Validation"). The CNBSS do not provide any information on women born before 1921 and after 1940, so that the birth cohort effect cannot be estimated in terms of the parameter ρ from this data set beyond 1921–40. To surmount this difficulty, the parameters ρ for other birth cohorts were calculated by using the rate ratios that resulted from the age–cohort model *(1,41–43)*. In doing so, we define ρ_1 associated with the 1921–25 cohort as a new baseline parameter while retaining the ratios ρ_2/ρ_1, ρ_3/ρ_1, and ρ_4/ρ_1 suggested by the analysis of the CNBSS data. The corresponding ratios for other birth cohorts are given by the analysis based on the age–cohort model. All birth cohorts were grouped in 5-year intervals and the estimated rate ratios were applied to the midpoints of these intervals. The estimated values of the parameter ρ for the various birth cohorts are shown in Fig. 1.

Remark 3. By no means can the age–period–cohort model replace or be superior to mechanistically motivated models, even when modeling cancer incidence in the absence of screening, because its structure is rather rigid, being completely determined by the assumption of proportionality of risks (rates). Besides, the relative risk variance tends to increase in calendar time due to an increasing extent of truncation of the baseline rate for late cohorts to eliminate the effect of screening.

A Simulation Model

Although many characteristics of the above-described model of the natural history of breast cancer can be derived analytically, we developed its simulation counterpart on which to explore the behavior of the basic model under various theoretical scenarios. This simulation model is easier to handle when comparing modeling results with epidemiological indicators in population

Table 2. Maximum likelihood estimates of model parameters with asymptotic 95% confidence intervals

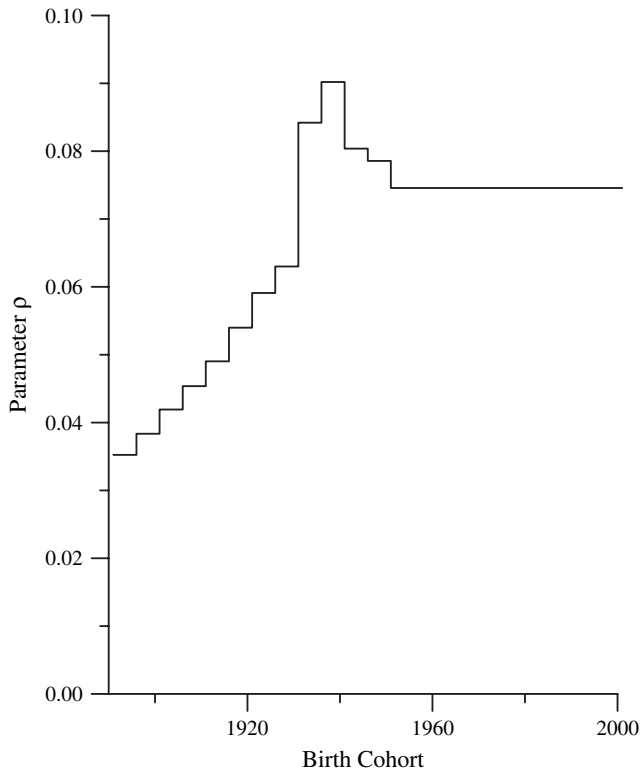| $\hat{A}$	$\hat{B}$	$\hat{\mu}$	$\hat{\sigma}$
1.112×10^{-4} (1.066×10^{-4}, 1.158×10^{-4})	0.1203 (0.1192, 0.1214)	0.526 (0.514, 0.537)	0.531 (0.489, 0.574)

Fig. 1. Estimated values of the parameter ρ as a function of birth cohort.

settings. Another advantage of the simulation approach is that the software can be more readily modified when new elements, such as sensitivity thresholds, need to be incorporated into the basic model structure. Also, the simulation model makes it easier to calculate such important characteristics as the mean lead time (and the corresponding variance) and program sensitivity. The simulation model generates individual histories of cancer development and detection for each birth cohort in accordance with the postulates formulated in "A Stochastic Model of Tumor Latency" and "Modeling the Impact of Screening on the Natural History of Breast Cancer". The time of tumor onset was generated according to the distribution given by formula [10.1], whereas for the preclinical stage duration W_0 the Gompertz distribution given by the second formula in [10.2] was used. The reciprocal of the growth rate was generated from a two-parameter gamma distribution. The effect of screening was modeled as described in "Modeling the Impact of Screening on the Natural History of Breast Cancer" (see "Mammographic Screening" for more details) with the probabilities of detection at the kth screen specified by formula [10.5]. The information on age and tumor size was retrieved after each event of either screen-based or spontaneous detection. The probabilistic characteristics of interest were estimated nonparametrically from the simulated data. The code was written in *PASCAL DELPHI*.

Breast cancer incidence. It is important to make inferences in terms of a characteristic of the model that can be estimated in the presence of data censoring. In a cohort setting, the most natural characteristic to be modeled is the hazard rate $h(x)$ as a function of age x at cancer detection. Under the model of independent censoring, the function $h(x)$ can be estimated from real or simulated data so that the resultant estimate does not depend on competing mortality.

Let $U_j = [x_{j-1}, x_j)$, then the life-table type estimator of $h(x)$ is given by

$$\hat{h}(x_j) = \frac{\text{number of events in U}_j}{\text{number at risk by the start of U}_j}, \quad [10.9]$$

so that there is no need to model competing risks explicitly. This is a distinct advantage of this indicator, because invoking independent information on competing mortality would induce an additional random noise in the epidemiological characteristic to be estimated. The estimator for $h(x)$ has desirable asymptotic properties: It is consistent and efficient. The same estimator can be used for mortality.

In a population setting, the hazard rate becomes time dependent and needs to be generalized leading to the notion of composite hazard. Let $h_i(x)$ be the hazard function for the ith cohort and t be the calendar year. The composite hazard $h^C(x, t)$ is defined as

$$h^C(x, t) = h_{t-x}(x). \quad [10.10]$$

Therefore, a pertinent estimator for $h^C(x, t)$ is

$$\hat{h}^C(x, t) = \hat{h}_{t-x}(x). \quad [10.11]$$

The empirical counterpart of $h^C(x, t)$ is

$$I(x_j, t) = \frac{\text{number of events (cases) in U}_j \text{ at time t}}{\text{number at risk by the start of U}_j \text{ at time t}}. \quad [10.12]$$

The commonly used indicator (age-specific incidence) is calculated as

$$I^*(x_j, t) = \frac{\text{number of new cases in U}_j \text{ at time t}}{\text{number of alive by the start of U}_j \text{ at time t}}. \quad [10.13]$$

In addition to the risk set, the denominator of [10.13] counts those persons in the age group U_j who have been diagnosed with cancer but are still alive in calendar year t.

The estimator I^* depends on the effects of data censoring (competing mortality), and there is no meaningful probabilistic characteristic for which the statistic $I^*(x_j, t)$ could be an unbiased estimator. If one uses $I^*(x_j, t)$ as an estimator for $h^C(x, t)$, the bias remains unknown. However, formulas [10.12] and [10.13] are expected to be numerically close to each other, and for this reason we believe that for all practical purposes the estimator $I(x_j, t)$ is well approximated by $I^*(x_j, t)$. For model calibration and validation, we use the incidence $I^*(x_j, t)$ and its age-adjusted (to the 2000 U.S. standard population) counterpart as meaningful summary characteristics of the SEER data. The age-adjusted true incidence is defined as

$$r(t) = \int h^C(x, t)\omega_0(x)dx, \quad [10.14]$$

where $\omega_0(x)$ is the age distribution in the standard population. When estimating r we replace h^C with I^*.

A model of breast cancer survival. To model mortality rates, we proceed from the following regression model *(13–15,17)* that relates the survival function, $\bar{G}$, of the postdiagnosis survival time to the values of clinical covariates (age, stage, tumor size) represented by vector $\mathbf{z}$:

$$\bar{G}(t \mid \boldsymbol{\beta}, \mathbf{z}) = \exp\left[-\theta(\boldsymbol{\beta}_1, \mathbf{z})\left\{ 1 - \bar{F}(t)^{\eta(\beta_2, \mathbf{z})} \right\} \right], \quad [10.15]$$

where $\underline{\boldsymbol{\beta}} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)$, $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of regression coefficients, $\bar{F}$ is an arbitrary survival function, and the functions θ and η are each of the form $\exp(\boldsymbol{\beta}'\mathbf{z})$.

Formula [10.15] is a natural generalization of the proportional hazards (PH) model with cure *(13,15)*; the latter is a special case of [10.15] with $\eta = 1$. A distinct advantage of this model is that each covariate may exert its effect both on long-term survival through θ (**z**) and on short-term survival through $\eta(\mathbf{z})$; this effect explains its higher flexibility compared with that of the traditional PH model. The need for extension [10.15] of the PH model is motivated by the fact that the original PH model does not provide a good description of breast cancer survival *(9,13,16,44,47)*. Within a semiparametric framework, the baseline function $\bar{F}$ is treated as a step function (with jumps at the observed failure times) which is set at zero at the point of last observation. Efficient algorithms are available to fit the semiparametric model [10.15] to survival data *(13–15)*. The model [10.15] has proven to provide an excellent fit to data on breast cancer *(13,16,47)* and prostate cancer *(17)* survival. The regression coefficients incorporated into θ (**z**) and η (**z**) were estimated from the SEER data by using an algorithm proposed by Tsodikov *(13)*; their numerical values are given in Table 3. In this analysis, we used survival data on more than 165 000 patients diagnosed with breast cancer since 1988. This subset was chosen because it provides the information on tumor size at diagnosis needed for our analysis. Similar estimates of the regression coefficients were obtained when the baseline function F was approximated by a two-parameter Weibull distribution.

In the simulation counterpart of our model, we generated a random variable, M, from the conditional survival function [10.15] for each set of covariates produced by the model of cancer detection (age, tumor size, clinical stage), with parameter values estimated from the Canadian studies (after a pertinent calibration of the model), so that the lifetime of each individual is equal to $U + M$. We did not analyze the CNBSS survival data because of their scarcity. The basic probabilistic characteristics of breast cancer mortality (such as the hazard rate) were estimated nonparametrically from the sample of simulated times $U + M$.

Mammographic screening. Although the model of breast cancer screening was described in sufficient detail earlier, a few further comments are in order here. To specify the initial value of the sensitivity parameter α_1 for the base case, we proceeded from its estimate obtained from the CNBSS data on the combined mode of detection, i.e., mammography and physical exam, because in real practice the two medical procedures frequently come together. To make the model of screening more realistic, we introduced threshold values for detectable volumes of tumors. The threshold volume for screening based detection was set at 0.004 cm^3, which is the minimum volume observed among screen-detected tumors in the CNBSS dataset. Similarly, a threshold of 0.014 cm^3 for spontaneous detection was determined from the CNBSS data after eliminating four smallest values suspected as likely outliers. However, the net results of modeling epidemiological descriptors are not perceptibly affected by the above thresholds.

Table 3. Regression coefficients estimated from the SEER data on breast cancer survival*

Covariate	Coefficient for θ (**z**)	Coefficient for η (**z**)
Baseline	$\beta_{11} = -2.11$	$\beta_{21} = 0$
Tumor size	$\beta_{12} = 3.74 \times 10^{-4}$	$\beta_{22} = 6.27 \times 10^{-4}$
Age at diagnosis	$\beta_{13} = 5.16 \times 10^{-6}$	$\beta_{23} = 5.33 \times 10^{-4}$
Stage, regional	$\beta_{14} = 1.30$	$\beta_{24} = 0.41$
Stage, distant	$\beta_{15} = 2.38$	$\beta_{25} = 1.18$

*SEER = Surveillance, Epidemiology, and End Results.

Individual schedules of mammographic examinations were modeled using the dissemination model developed by the NCI. This software generates a screening schedule for each individual pertaining to a given birth cohort. In addition to this sequence of screening ages, each individual history of breast cancer includes random variables T and Λ, as well as the times W_0 and W_1 of spontaneous and screen-based detection, respectively. Both random variables W_0 and W_1 are measured from the time of tumor onset.

Given T, Λ, and an individual screening schedule $\tau_1, \tau_2, \ldots, \tau_n$, the random variables W_0 and W_1 are generated using the second of formulas [10.2] and formula [10.5], respectively, which gives a sample value of $W = \min(W_0, W_1)$. The components W_0, W_1 determining the actual age at detection are only conditionally independent, given the time T of tumor onset. Therefore, these components cannot be manipulated independently to achieve a better fit to the observed data.

Once the age at tumor detection $U = T + W$ has been determined, a check is made as to whether its value exceeds the maximum allowable age in a given cohort. If it does not, the size of the detected tumor is recorded. Thus, the output of our simulations is represented by the pairs U, S. The clinical stage (local, regional, distant) is generated conditionally on this output from a distribution estimated from the SEER data, yielding triples of quantities that are necessary to construct the most basic epidemiological indicators.

Modeling Effects of Treatment

As described earlier in this report, the effect of early detection on mortality was modeled through the regression coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ characterizing the contributions of age, tumor size, and clinical stage to short- and long-term survival effects, respectively. Maximum likelihood estimates of these coefficients were obtained from the SEER data on postdetection survival of patients with breast cancer diagnosed after 1988. This time interval is characterized by a widespread use of novel modes of adjuvant therapy for breast cancer, first and foremost of those associated with the advent of tamoxifen. When modeling the base case, however, one needs to cover the whole interval between 1975 and 1999. Therefore, using the coefficients $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ thus estimated would result in a significantly lower mortality than that was actually observed. The SEER data do not provide the necessary information on breast cancer treatment so that the effect of tamoxifen and other advancements in breast cancer treatment has to be modeled by the indirect route.

One way of doing this is to calibrate the model by introducing two additional time-dependent covariates z_θ and z_η and the corresponding scaling parameters $\exp(c_\theta z_\theta)$ and $\exp(c_\eta z_\eta)$ that modify the short- and long-term survival effects by acting multiplicatively on the functions $\theta(\boldsymbol{\beta}_1, \mathbf{z}) = \exp(\boldsymbol{\beta}'_1 \mathbf{z})$ and $\eta(\boldsymbol{\beta}_2, \mathbf{z}) = \exp(\boldsymbol{\beta}'_2 \mathbf{z})$ in formula [10.15]. The effect of treatment on breast cancer mortality needs to be modeled as a function of calendar time, t, to reflect the dissemination of tamoxifen and other therapy improvements. To retain identifiability of the model, we assume that there is a change point t_0 (calendar year) so that $z_\theta = 1$, $z_\eta = 1$ for $t < t_0$ and $z_\theta = z_\eta = 0$ for $t \geq t_0$. Thus, we introduce the simplest stepwise dependence of the effect of treatment on calendar time. This model will be referred to as Model 1. This gives us three more parameters c_θ, c_η and t_0 to calibrate the model of breast cancer mortality. The rationale for model calibration is

discussed in the next section. Using the SEER data we obtained the following least squares estimates: $c_\theta = 2.65$, $c_\eta = -3.05$ and $t_0 = 1980$. These values provide a reasonably good fit to the observed breast cancer mortality over 1975–1999 (Fig. 4), and at the same time they serve as auxiliary quantitative characteristics of the contribution of therapy advancements (including tamoxifen) to breast cancer survival. In particular, extending the estimated values of c_θ and c_η to the period $t \geq t_0$ would yield a mortality rate that would be observed with no use of tamoxifen (no improvements in treatment), while setting $c_\theta = c_\eta = 0$ for all t one could predict a mortality rate that would have been observed had the contemporary modes of treatment been in effect since the beginning of the twentieth century.

A more realistic description of the observed mortality trend can be provided by introducing a gradual advent of improved treatments (better surgical procedures, improved irradiation regimens, adjuvant chemotherapy, patient care, etc) that has begun before 1975. A simple model (Model 2) is derived by assuming that both $z_\theta(t)$ and $z_\eta(t)$ are linearly decreasing functions such that $z_\theta(t_1) = z_\eta(t_1) = 1$ and $z_\theta(t_2) = z_\eta(t_2) = 0$, where $t_1 < 1975$ and $t_2 > 1975$. As shown in Fig. 5, a nearly perfect fit to the observed breast cancer mortality is provided by this model with $t_1 = 1960$ and $t_2 = 1990$.

MODEL VALIDATION

General Principles

Assessing goodness of fit for the model described above is difficult because of its complex and multivariate structure. There are no theoretically based statistical methods of goodness-of-fit testing for the bivariate distribution given by formula [10.4], whereas resampling and cross-validation techniques are computationally prohibitive with a model of such complexity. Data censoring and truncation also stand in the way as far as the CNBSS data are concerned. The CNBSS data are heterogeneous with respect to individual screening schedules, which is why nonparametric estimators of such important quantities as the distribution of tumor size at detection or its mean value may be biased in finite samples, thereby causing more complications in goodness-of-fit testing.

For all these reasons we use the base case only to validate the model. In doing so, the CNBSS data will serve as a training set, whereas the SEER data will be treated as a control sample. This validation design is typical of supervised learning methods. Unlike situations in discriminant analysis, where outcome variables are categorical, we have to compare two continuous functions representing parametric (model based) and nonparametric estimates of the epidemiological indicator of interest. Statistical goodness-of-fit tests are of little utility in comparing the expected values predicted by the model with the observed values of epidemiological indicators in the base case from the following considerations:

1. In large-sample studies, goodness-of-fit tests may be overly conservative rejecting any reasonable (no model is perfect) model.
2. Even if one is prepared to assume the Poisson error structure [which is not a plausible hypothesis in the presence of screening (2)], it is still extremely difficult to make use of asymptotic results for the sampling distribution of a statistic based on residuals, because the parameters are not estimated from the same data. For example, the asymptotic

sampling distribution of the chi-square statistic becomes complicated when a distribution with parameters estimated from one set of data is tested for goodness of fit to some other set (45).

Therefore, we rely on graphical methods based on residuals characterizing the discrepancy between the observed population-based indicators (rates) and their values predicted by the best-fit model. When estimating model parameters from a given set of data, there is always a danger of overfitting, that is, fitting overly specific patterns that do not extend to new samples. This kind of overfitting has to do with model flexibility; it may manifest itself even if a model is identifiable and all its parameters are properly estimated [see (46) for discussion of the difference between the explained variation and predictive properties of a model]. The phenomenon of overfitting is also known in regression analysis as the shrinkage effect—which is why the model needs to be calibrated when tested against the control sample. Calibration should not end up with the reestimation of all parameters from the control data set; otherwise, no conclusion regarding predictive qualities can be made. In other words, a calibration procedure should be as parsimonious as possible. We require also that at least some predictions be made with no further calibration of the model.

Calibration of Model

There are two principles of parsimony we tried to follow in this work:

1. Calibration may be applied to a given parameter if there are biological grounds to believe that this parameter may indeed vary between the two settings under comparison (e.g., variations in risk factors, sensitivity of screening procedures). A calibration procedure may also involve those parameters that cannot be estimated from the training set for the lack of relevant data.
2. The number of parameters involved in calibration should be kept to a minimum.

In our calibration procedure, the mean growth rate μ was fixed at its value of 0.526 obtained from the CNBSS data. However, we included σ in the procedure, because we expected more heterogeneity in the population-based SEER data than in the CNBSS dataset generated by controlled screening trials. The sensitivity parameters α_0 and α_1 may also vary between the two sets of data.

To meet the second requirement, we can take advantage of some properties of the model described below. These properties have to do with a relatively low sensitivity of some epidemiological indicators to a certain subset of parameters. To calibrate the model we used the so-called incidence size distribution defined as follows. Let $r_j(t)$ be the age-adjusted incidence for the jth tumor size category (range), $j = 1, \ldots, k$, then the incidence size distribution at time t is given by

$$\varphi(j, t) = \frac{r_j(t)}{\sum_{i=1}^{k} r_i(t)}, \qquad [10.16]$$

where t is calendar time. Our calibration procedure involves the following steps:

Step 1. Since the distribution $\varphi(j, t)$ for $t = 1975$ is practically insensitive to the parameters ρ_i characterizing the birth cohort effect (this conjecture was corroborated by computer

simulations) and the effect of screening (reflected in the parameter α_1) is expected to be negligibly small in 1975, we fit the model to the observed size distribution (three size categories) by minimizing the sum of squared residuals with respect to only two parameters: α_0 and σ, while setting $\alpha_1 = 0$.

Step 2. In 1999, we expect the size distribution to depend predominantly on α_1. Therefore, we fit this distribution by the method of least squares, changing only α_1 and setting the parameters α_0 and σ at their values resulted from Step 1.

Step 3. We repeat Step 1 with the newly estimated α_1 and then proceed to Step 2.

We alternate between the first two steps until a stable solution is obtained; just two iterations are normally needed to obtain such a solution. Clearly this algorithm can be improved by sequentially including more time points in the objective function when alternating between the two steps. In our preliminary studies, we used only the simplest version of the algorithm.

Remark 4. The model-based marginal distribution of tumor size evaluated at a given time point (say, at $t = 1975$) is no longer a Pareto distribution even in the absence of screening, because it involves the condition that the age at detection does not exceed a certain value. This is all the more so for the distribution $\varphi(j, t)$. Therefore, it is not recommended to use the Pareto approximation in Step 1 of the above algorithm.

The above calibration procedure was applied to the SEER data on invasive breast cancer (excluding all in situ tumors) with all ages included in the adjustment of the age-specific incidence to the 2000 U.S. standard population. The resultant estimate $\sigma = 0.60$ indicates that the distribution of tumor growth rate may be slightly overdispersed. The estimate of σ is slightly larger than the maximum likelihood estimate of $\hat{\sigma} = 0.53$ from the CNBSS data; the observed tendency is consistent with the fact that the SEER data are more heterogeneous than the CNBSS data. The sensitivity parameters α_0 and α_1 were estimated as 4.48×10^{-10} and 8.30×10^{-7}, respectively. It is just natural that the calibrated parameter α_0 tends to be slightly smaller than its maximum likelihood estimate obtained from the CNBSS, because all participants in the latter study received self-examination instruction. However, a much higher value of α_1 still awaits interpretation (see "Discussion"). The comparison of the size distributions resulted from this procedure and their empirical counterparts is shown in Table 4.

The initial values of ρ_i (initiation rates) for each cohort were obtained by our analysis of the CNBSS data on age and tumor size at detection followed by the application of the age–cohort model. Almost no calibration ($K_\rho = 1.02$, see below for definition) was necessary of the size-specific incidence (for tumors of known size) with respect to these parameters. Therefore, the
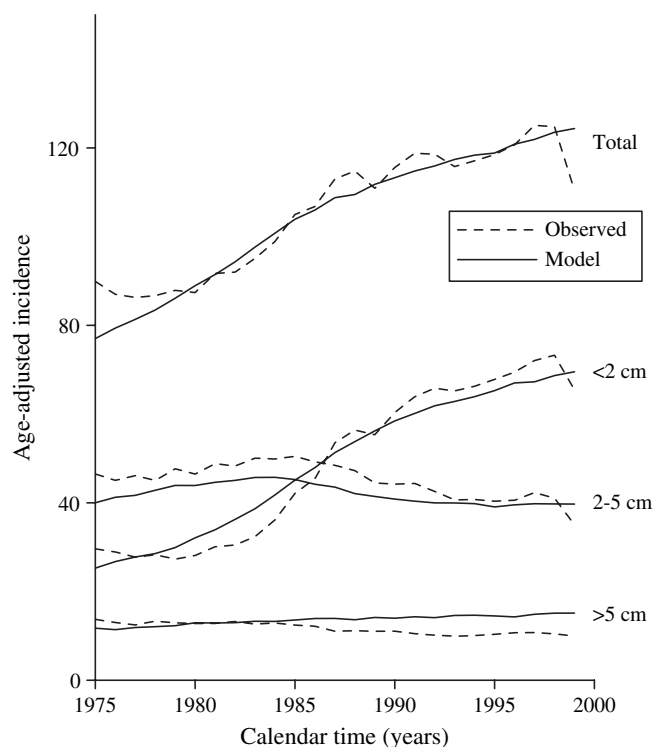

Fig. 2. Predicting size-specific age-adjusted (all ages) breast cancer incidence at fixed parameters of the model. Only invasive tumors of known size are included.

results of our predictions pertaining to the size-specific incidence (see "Predictive Properties") were effectively obtained using the maximum likelihood estimates of ρ_i for the four birth cohorts in the CNBSS data and the relative risks estimated under the age–cohort model.

However, the situation is not the same for the total age-adjusted incidence that includes counts of tumors with missing size information. To predict this epidemiological descriptor, an additional calibration of the model with respect to ρ_i is absolutely necessary. This process amounts to imputation of missing data on the number of cases with unknown tumor sizes. Indeed, it is impossible for a model based on tumor size at detection to provide a description of the contribution of cases with unknown sizes to the overall incidence, because the model requires that the total age-adjusted incidence be equal to the sum of size-specific age-adjusted incidence curves. To keep the extent of this additional calibration to a minimum, all parameters ρ_i were multiplied by the same (independent of i) scaling factor, K_ρ, chosen as a minimizer of the corresponding sum of squared residuals. This scaling procedure appears to have no tangible effect on the other parameters and on the quality of our predictions, so that no further tuning of the model is warranted. Thus, the parameter K_ρ plays essentially the same role as the shrinkage factor in the predictive regression analysis. For the total age-adjusted incidence we report $K_\rho = 1.14$, whose value was obtained by calibration of the model to fit the observed incidence that includes cases with unknown tumor size. The total age-adjusted incidence is not the only example where the calibration with respect to ρ_i may be required to compensate for missing information; there may be other cases (where the results of modeling are extrapolated to another data set) calling for such a calibration.

Table 4. Model fit to the incidence size distribution*

Tumor size (diameter, cm)	1975		1999	
	Observed (%) (SEER)	Model (%)	Observed (%) (SEER)	Model (%)
<2	32.94	32.81	59.24	55.05
2–4.9	51.73	52.23	31.75	32.74
≥5	15.27	14.96	9.01	12.21

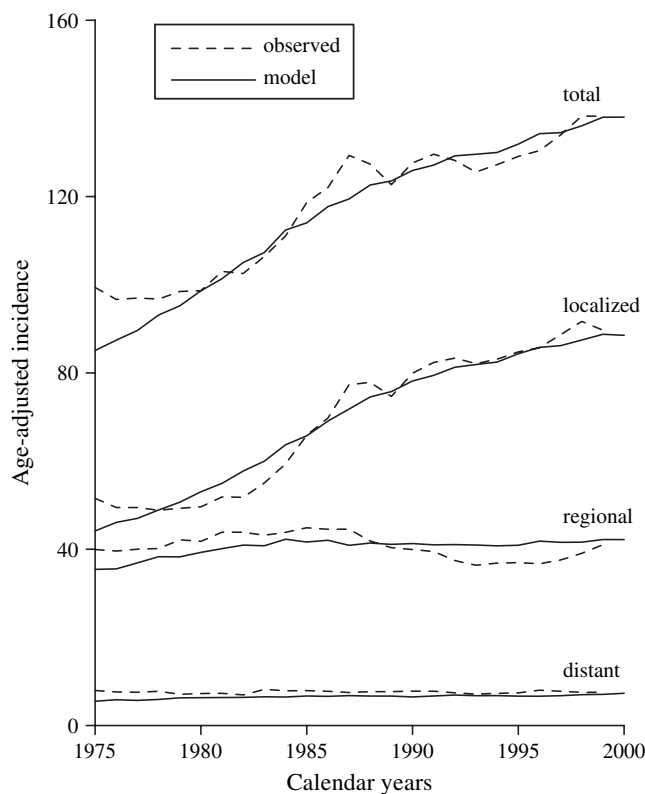*SEER = Surveillance, Epidemiology, and End Results.

Fig. 3. Predicting stage-specific age-adjusted (all ages) breast cancer incidence. Model predictions (**solid lines**); observed incidence curves (**dashed lines**). Only invasive tumors of known stage are included.

Fig. 4. Predicting age-specific breast cancer incidence at fixed parameters of the model. Only invasive tumors of known size are included. The same notation as in Figs. 2 and 3 is used.

Predictive Properties

Now we can validate the model by predicting (with no further calibration or tuning) certain quantitative characteristics obtained from the SEER data. In particular, we would like to predict the dynamics of the following indicators:

1. Size-specific (three size categories) and stage-specific age-adjusted (all ages) incidence curves as functions of calendar time as well as the total (excluding tumors of unknown size) age-adjusted incidence of malignant breast cancer over 1975–1999.
2. Age-specific incidence for cases of invasive breast cancer with known tumor sizes.

As is obvious from Figure 2, the model well describes the size-specific age-adjusted breast cancer incidence at fixed values of all parameters. Shown in Fig. 3 are the stage-specific (three stages) age-adjusted (all ages) incidence and the total (excluding unstaged tumors) age-adjusted incidence of malignant breast cancer. The mechanism generating missing data on tumor size is not purely random and appears to depend on calendar time, which is why we need to identify and eliminate such cases from the SEER data rather than attempting to model this mechanism. Fig. 4 shows sample predictions of the age-specific incidence (cases with known tumor size) which appear to be surprisingly good, given that the parameters α_0 and α_1 were held constant across all age groups and none of these curves was used for calibration; *the model was calibrated only to the incidence size distribution at two time points*. The results for other years (not shown because of space limitations)
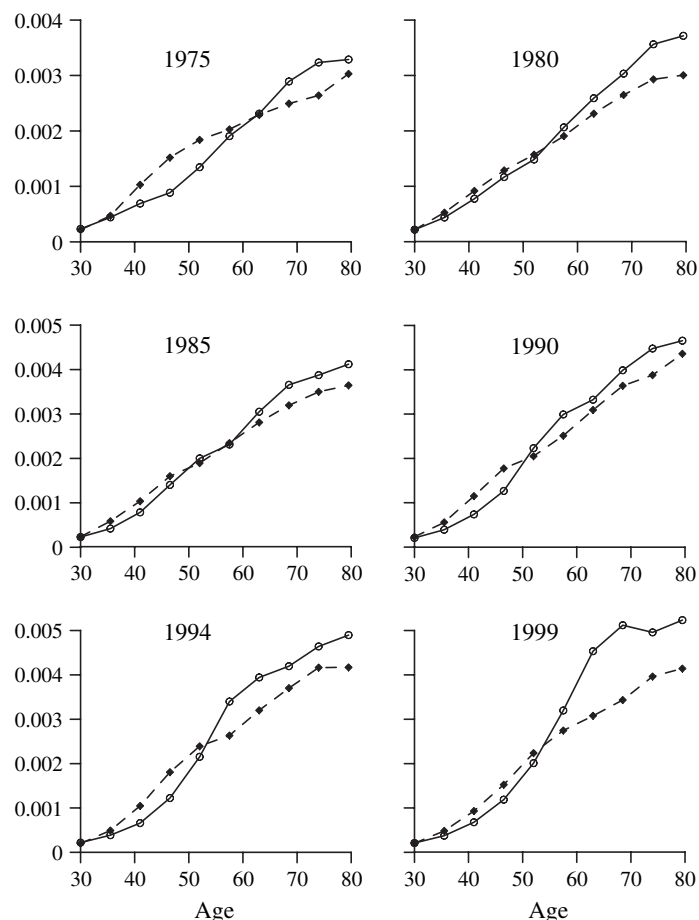
are similar. The only notable discrepancy observed in 1999 is somehow related to the fact that the age-adjusted incidence displays some sort of irregular behavior in the vicinity of this time point (see Fig. 2).

Figure 5 shows how the model fits the observed breast cancer mortality. Although it is clear that breast cancer incidence continues to increase after 1975, the mortality curve is flat for a period of 15 years. It is impossible to explain these trends by screening by treatment interactions in view of the fact that such an effect may show up only after a time delay. In contrast, incremental improvements in therapy (before and after 1975) provide a likely explanation. It is seen from Fig. 5 that Model 2 improves the fit dramatically in comparison to Model 1 as far as the early portion of the mortality curve is concerned. Recall that Model 1 assumes a stepwise change in treatment efficacy occurring at some time point t_0, while a more gradual (linear) trend is incorporated into Model 2. As is obvious from Fig. 5, the effects of screening by treatment interactions begin manifesting themselves in mortality after 1990. These results clearly demonstrate that the model captures the most salient features of the processes under study.

DISCUSSION

We begin our discussion by quoting Clayton and Schifflers *(1)*: It is the purpose of statistical analysis to extract from research data the maximum information in as parsimonious and
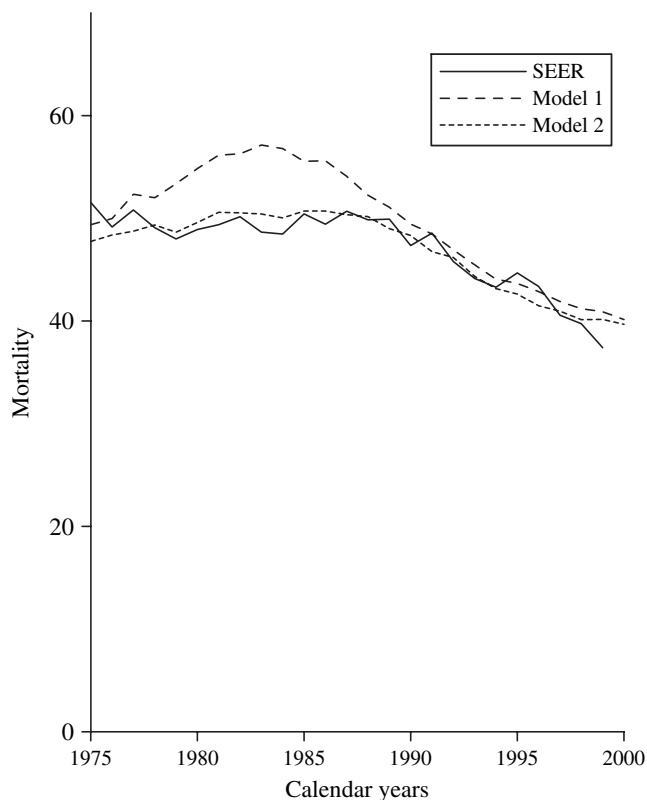
Fig. 5. Age-adjusted (30–79 years) breast cancer mortality for the period between 1975 and 1999. SEER = Surveillance, Epidemiology, and End Results.

Fig. 6. Predicting breast cancer mortality (Model 2) in the absence or presence of screening.

comprehensive manner as possible. Although absolutely valid, this statement places two conflicting requirements upon model-based statistical inference. For a model to be useful, its complexity must be adequate for the information contained in the data to be analyzed. A mathematical or simulation model whose parameters are not identifiable is of no use for data analysis, unless a proper reparameterization results in identifiable combinations of model parameters. If such combinations cannot be found, more sources of information need to be used to overcome this difficulty. Wherever possible, a theoretical proof of model identifiability should be provided. Alternatively, numerical or simulation studies are needed to show that the model is not overparameterized and is sensitive enough to parameter values to allow for estimation of its parameters from real data. A model must be sufficiently simple to meet the above requirements. At the same time, it should be flexible enough to provide a good description of heterogeneous data sets. The approach presented here appears to satisfy both requirements, thereby representing the desired compromise between identifiability and flexibility of the proposed model.

We use a fully parametric model of the natural history of cancer for making maximum likelihood inferences from randomized screening trials. Having estimated parameters of the model from such data, one can use a simulation counterpart of the same model to predict various indicators associated with breast cancer incidence and mortality in the general population under different screening scenarios. In predictive settings, where model parameters are estimated from some other dataset, calibration is necessary to mitigate the effect of overfitting. This is an important step in an attempt to extrapolate the initial parametric inference from a randomized trial to a population-based setting. The proposed
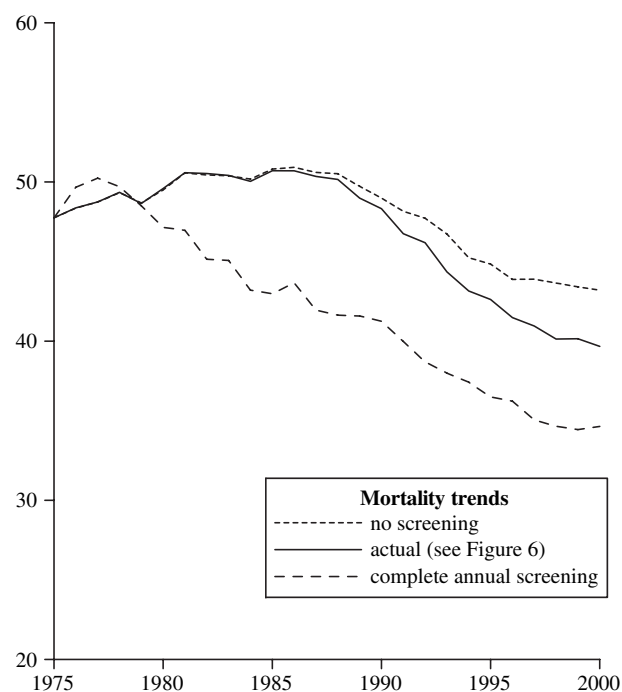
model structure is well suited to calibrate the model in a parsimonious and biologically meaningful way. This goal is accomplished through designing a stepwise fitting procedure so that the parameters α_0 and σ are chosen to fit the tumor size distribution observed when the dissemination of mammography is believed to be low, while the parameter α_1 is estimated to fit the same distribution at the end of the observation period. Relative insensitivity of the size distribution to certain subsets of parameters helps design such a procedure. The calibration procedure thus designed can also be viewed as a method for estimating some parameters of the natural history from data on cancer incidence in the general population. For example, the mean growth rate μ can also be estimated from the incidence size distribution (Step 2 of the proposed algorithm), which may improve the fit shown in Table 4 for $t = 1999$. Just to make our validation procedure as strict as possible, we intentionally refrained from adjusting the parameter μ. However, estimation of the parameters incorporated into the onset time distribution, like ρ, A, and B, calls for cohort observations. Randomized trials represent the best designed cohort studies, which is why we combine both types of parametric inference in the analysis of cancer incidence.

The value of α_1 estimated from the SEER data appears to be much higher than its maximum likelihood estimate obtained from the Canadian study. This discrepancy can be attributed to the fact that the CNBSS data include in situ tumors, while the calibrated parameter α_1 refers to invasive breast cancer. Yet another possibility is that the NCI model of mammography dissemination may underestimate the actual intensity of screening, so that the model compensates for this bias yielding a higher value of the sensitivity parameter α_1. The latter explanation is speculative, of course.

The model shows an excellent description of the observed breast cancer incidence and mortality in the U.S. population. The mortality trend is consistent with the assumption that there has

been a relatively long history of incremental improvements in breast cancer treatment. There are two reasons why we refrained from using the results of meta-analysis based on the proportional hazards model to explicitly describe the effect of adjuvant chemotherapy on breast cancer mortality. First, we know that the Cox model does not provide a good description of covariate effects on breast cancer survival *(9,13,16,44,47)*. Second, one can see that the age-adjusted mortality curve is flat for 15 years (beginning from 1975), whereas breast cancer incidence continues to increase. This pattern indicates that improvements in breast cancer treatment began manifesting themselves before the start of any appreciable dissemination of mammography in the U.S. population. A similar observation was recently reported for breast cancer mortality in the United Kingdom *(18)*.

The contribution of screening to the observed decline in mortality appears to be rather weak under our model (Fig. 6). The main point here is that the actual dissemination of screening in the U.S. population is too low for a tangible survival benefit from mammography due to screening by treatment interactions. It is also quite low in screening trials because of a narrow range of screening ages and just a few scheduled examinations. If the survival benefit of screening in randomized trials were truly strong it would inevitably be seen far beyond the well-known breast cancer screening controversy *(48)*. To appreciate such a benefit, the target population must be subjected to a much more intensive screening. To demonstrate this, we ran the model in a way that mimics annual screening of all women older than 30 over 1975–1999 (special run). In this run, the percent decline in mortality due to screening is expected to be more than 19.8% by 1999 (Fig. 6). Thus, the model indicates a significant benefit of breast cancer screening providing its dissemination is sufficiently intensive. The estimated mean lead time is 2.06 years (all ages) and does not change much in the special run.

However nice a final fit may be, it is not enough for model validation. One needs to evaluate predictive properties of the model under study in a situation where no further calibration is allowed. In some predictive settings (e.g., where missing information is included in the indicator to be predicted; see "Calibration of the Model") there is no way to obviate the need for an additional calibration, although such situations should be avoided whenever possible. If such a calibration appears to be unavoidable, it should involve as few parameters as possible.

We failed to explain the observed increase of breast cancer incidence by mammography dissemination alone. At no reasonable parameter values does the model fit the data after removing the birth cohort effect. This shows that the model is realistic enough to reject unrealistic scenarios.

Appendix: Basic Notation

T - age of an individual at tumor onset;

A and B - parameters incorporated into the distribution of T;

ρ - ratio of the initiation rate and the rate of proliferation of initiated cells;

ρ_i - parameter ρ for the ith birth cohort;

W - time interval between T and the age at tumor detection;

U - $U = T + W$;

S - tumor size at detection;

λ - rate of exponential tumor growth;

Λ - random rate of tumor growth;

μ - expected value of $1/\Lambda$;

σ - standard deviation of $1/\Lambda$;

α_0 - sensitivity parameter (proportionality coefficient in a quantal response model) for clinical detection;

α_1 - sensitivity parameter (proportionality coefficient in a quantal response model) for mammography + physical exam;

α_2 - sensitivity parameter (proportionality coefficient in a quantal response model) for physical examination alone;

M - postdetection survival time;

$G(\cdot)$ - survival time cumulative distribution function;

$\bar{G}(\cdot)$ - survival function: $\bar{G}(\cdot) = 1 - G(\cdot)$;

z - vector of covariates;

β - vector of regression coefficients;

t_0 - change point in calendar time;

c_θ, c_η - calibration coefficients.

References

(1) Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. Statistics in Medicine 1987b;6:469–71.

(2) Hanin LG, Yakovlev AY. Multivariate distributions of clinical covariates at the time of cancer detection. Stat Methods Med Res 2004;13:457–89.

(3) Zelen, M. A hypothesis for the natural time history of breast cancer. Cancer Research 1968;28:207–16.

(4) Feldstein M, Zelen M. Inferring the natural time history of breast cancer: implications for tumor growth rate and early detection. Breast Cancer Res Treat 1984;4:3–10.

(5) Blumenson LE, Bross ID. A mathematical analysis of the growth and spread of breast cancer. Biometrics 1969;22:95–109.

(6) Schwartz M. An analysis of the benefits of serial screening for breast cancer based upon a mathematical model of the disease. Cancer 1978;41:1550–64.

(7) Schwartz M. A mathematical model used to analyse breast cancer screening strategies. Oper Res 1978;26:937–55.

(8) Baker SG, Erwin D, Kramer BS, Prorok PC. Using observational data to estimate an upper bound on the reduction in cancer mortality due to periodic screening, BMC Med Res Methodol 2003;3:4. Available at: http://www.biomedcentral.com/1471-2288/3/4.

(9) Yakovlev AY, Tsodikov AD. Stochastic models of tumor latency and their biostatistical applications. Singapore: World Scientific; 1996.

(10) Asselain B, Fourquet A, Hoang T, Tsodikov AD, Yakovlev AY. A parametric regression model of tumor recurrence: an application to the analysis of clinical data on breast cancer. Stat Probabil Lett 1996;29:271–8.

(11) Ibrahim JG, Chen MH, Sinha D. Bayesian survival analysis. New York (NY): Springer; 2001.

(12) Ibrahim JG, Chen MH, Sinha D. Bayesian semi-parametric models for survival data with a cure fraction. Biometrics 2001;57:383–8.

(13) Tsodikov A. Semiparametric models of long- and short-term survival: an application to the analysis of breast cancer survival in Utah by age and stage. Stat Med 2002;21:895–920.

(14) Tsodikov A. Semiparametric models: a generalized self-consistency approach. J R Stat Soc Ser B 2003;65:759–74.

(15) Tsodikov AD, Ibrahim JG, Yakovlev AY. Estimating cure rates from survival data: an alternative to two-component mixture models. JASA 2003;98:1063–78.

(16) Yakovlev AY, Tsodikov AD, Boucher K, Kerber R. The shape of the hazard function in breast carcinoma: curability of the disease revisited. Cancer 1999;85;1789–98.

(17) Zaider M, Zelefsky MJ, Hanin LG, Tsodikov AD, Yakovlev AY, Leibel SA. A survival model for fractionated radiotherapy with an application to prostate cancer. Phys Med Biol 2001;46:2745–58.

(18) Kobayashi S. What caused the decline in breast cancer mortality in the United Kingdom? Breast Cancer 2004;11:156–9.

(19) Albert A, Gertman PM, Louis TA, Liu SI. Screening for the early detection of cancer. 2. The impact of the screening on the natural history of the disease. Math Biosci 1978;40:61–109.

(20) Bartoszyński R, Edler L, Hanin L, Kopp-Schneider A, Pavlova L, Tsodikov A, Zorin, A, Yakovlev A. Modeling cancer detection: tumor size as a source of information on unobservable stages of carcinogenesis. Math Biosci 2001;171:113–42.

(21) Moolgavkar SH, Venzon DJ. Two event model for carcinogenesis: Incidence curves for childhood and adult tumors. Math Biosci 1979;47: 55–77.

(22) Moolgavkar SH, Knudson AG. Mutation and cancer: a model for human carcinogenesis. J Natl Cancer Inst 1981;66:1037–52.

(23) Moolgavkar SH, Luebeck EG. Two-event model for carcinogenesis: Biological, mathematical and statistical considerations. Risk Anal 1990; 10:323–41.

(24) Heidenreich WF. On the parameters of the clonal expansion model. Radiat Environ Biophys 1996;35:127–9.

(25) Hanin LG, Yakovlev AY. A nonidentifiability aspect of the two-stage model of carcinogenesis. Risk Anal 1996;16:5:711–5.

(26) Heidenreich WF, Luebeck EG, Moolgavkar SH. Some properties of the hazard function of the two-mutation clonal expansion model. Risk Anal 1997;17:391–9.

(27) Gregori G, Hanin L, Luebeck G, Moolgavkar S, Yakovlev A. Testing goodness of fit with stochastic models of carcinogenesis. Math Biosci 2001;175:13–29.

(28) Hanin L. Identification problem for stochastic models with application to carcinogenesis, cancer detection and radiation biology. Discrete Dynamics Nat Soc 2002;7:177–89.

(29) Zorin AV, Edler L, Hanin LG, Yakovlev AY. Estimating the natural history of breast cancer from bivariate data on age and tumor size at diagnosis. In: Quantitative Methods for Cancer and Human Health Risk Assessment, L. Edler and C.P. Kitsos, editors. New York (NY): Wiley; 2005. pp. 317–27.

(30) Hanin LG, Yakovlev AY. Identifiability of the joint distribution of age and tumor size at detection in the presence of screening, Math Biosci. 2004, submitted.

(31) Mandelblatt J, Saha S, Teutsch S, Hoerger T, Siu AL, Atkins D, et al. A systematic review: the cost-effectiveness of screening mammography beyond age 65. Ann Intern Med 2003;139:835–42.

(32) Hanin LG, Tsodikov AD, Yakovlev AY. Optimal schedules of cancer surveillance and tumor size at Detection. Math Comput Model 2001;33: 1419–30.

(33) Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data. Springer Series in Statistics for Biology and Health. New York (NY): Springer; 1997.

(34) Miller AB, Howe GR, Wall C. The national study of breast cancer screening. Clin Invest Med 1981;4:227–58.

(35) Miller AB, Baines CJ, To T, Wall C. Canadian national breast screening study: 1. Breast cancer detection and death rates among women age 40–49 years. Can Med Assoc J 1992a;147:1459–76 (published erratum in Can Med Assoc J 1993;148:718).

(36) Miller AB, To T, Baines CJ, Wall C. The Canadian National Breast Screening Study—1. A randomized screening trial of mammography in women age 40–49: Breast cancer mortality after 11–16 years of follow-up. Ann Intern Med 2002;137:5:305–12.

(37) Miller AB, Baines CJ, To T, Wall C. Canadian National Breast Screening Study 2. Breast cancer detection and death rates among women aged 50 to 59 years. Can Med Assoc J 1992b;147:1477–88 (published erratum in Can Med Assoc J 1993;148:718).

(38) Miller AB, To T, Baines, CJ, Wall C. Canadian National Breast Screening Study 2: 13-year results of a randomized trial in women age 50–59 years. J Natl Cancer Inst 2002;92:1490–9.

(39) Pflug, G. C. Optimization of stochastic models: the interface between simulation and optimization. Boston (MA): Kluwer Academic Publishers; 1996.

(40) Boucher KM, Kerber RA. The shape of the hazard function for cancer incidence. Math Comput Model 2001;33:1361–76.

(41) Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. Stat Med 1987;6:449–67.

(42) Wun LM, Feuer EJ, Miller BA. Are increases in mammographic screening still a valid explanation for trends in breast cancer incidence in the United States? Cancer Causes Control 1995;6:135–44.

(43) Tarone RE, Chu KC. Age-period-cohort analyses of breast-, ovarian-, endometrial- and cervical-cancer mortality rates for Caucasian women in the USA. J Epidemiol Biostat 2000;5:221–31.

(44) Pocock SJ, Gore SM, Kerr GR. Long-term survival analysis: the curability of breast cancer. Stat Med 1982;1:93–104.

(45) Greenwood PE, Nikulin MS. A guide to chi-squared testing, New York (NY): Wiley Interscience; 1996.

(46) Verweij PJM, Van Houwelingen HC. Cross-validation in survival analysis. Stat Med 1993;12:2305–14.

(47) Boucher K, Asselain B, Tsodikov AD, Yakovlev AY. Semiparametric versus parametric regression analysis based on the bounded cumulative hazard model: an application to breast cancer recurrence. In: Nikulin M, Balakrishnan N, Mesbah M, Limnious N, eds., Semiparametric models and applications to reliability, survival analysis and quality of life. Birhäuser; 2004. pp. 399–418.

(48) Olsen O, Gotzsche PC. Cochrane review on screening for breast cancer with mammography. Lancet 2001;358:1340–2.

NOTES