Biometrics 60, 191–198
March 2004

An Early- and Late-Stage Convolution Model for Disease Natural History

Paul F. Pinsky

Division of Cancer Prevention, National Cancer Institute, 6130 Executive Blvd., EPN 3064, Bethesda, Maryland 20892, U.S.A.
email: pinskyp@mail.nih.gov

Summary. The standard convolution model of disease natural history posits an asymptomatic (preclinical) and a symptomatic (clinical) state. An augmented model includes, in both the preclinical and clinical states, an early and late stage of disease. In the case of cancer, the early stage would generally correspond to the organ-confined stages before there is evidence of cancer spread. We compute the number of screen-detected (preclinical) and clinical cases in the early and late stages expected under a given screening program and show how the model can be fit to data from a screening trial using maximum likelihood. We also develop expressions for sojourn time, lead time, and overdiagnosis in the context of the model, where each of the above concepts incorporates disease stage. As an example, we fit the model to data from the Mayo Lung Cancer Screening trial.

Key words: Cancer; Lead time; Natural history; Screening; Stage.

1. Introduction

Convolution models have frequently been used to describe the natural history of a disease of interest (Day and Walter, 1984; Paci and Duffy, 1991; Chen, Duffy, and Tabar, 1996). These models posit that disease incidence is the convolution of two densities, a density for progression into an asymptomatic (or preclinical) disease state and a density for progression from the asymptomatic to the symptomatic (clinical) state. Although the model parameters are not identifiable from clinical disease data alone, they are identifiable (in general) when subjects can be identified in the preclinical disease state. Early detection or screening for disease attempts to identify subjects in a preclinical disease state; thus data from screening studies have often been used for fitting convolution models.

Models with more than one preclinical disease state could have added utility as compared to the above standard convolution model with a single preclinical state. For example, in most cancers, the stage of disease is determined at (preclinical or clinical) diagnosis and is a strong prognostic factor for disease survival and response to treatment. As such, it is critical for a cancer model to distinguish between two (or more) stages of preclincial (and clinical) disease. A standard convolution model can be used to calculate how many cases would be seen in the preclinical state during a particular screening program; a model with an early and late preclinical state could be used to calculate not only how many of these cases would be early- versus late-stage cancers, but also how many of those with screen-diagnosed early-stage cancers would have progressed to clinical late-stage cancer in the absence of screening.

Chen et al. (2000) recently described a five-state Markov natural history model for breast cancer that included two preclinical states, essentially an early- and late-stage preclinical state, along with an early- and late-stage clinical state (and a normal state). They attempted to fit this model to screening data without using information on clinical cases. They concluded that a straight likelihood approach was not possible without the clinical cases and used a reparameterization procedure in conjunction with external information to fit the model.

Here, we utilize the basic framework of the model described by Chen et al. (2000). However, we add competing mortality and generalize the model so that it no longer needs to be Markov, but instead is a general convolution model. Relaxing the Markov assumption is especially important when modeling the transition from the normal to the (early) preclinical state, which would be expected to vary markedly with age. We demonstrate how this model can be fit to screening trial data, which includes clinical as well as screen-detected cases, using maximum likelihood. In addition, we rigorously define, in the context of the current model and incorporating disease stage, various quantities such as sojourn time, lead time, and overdiagnosis, which are often utilized to help judge the potential impact of screening programs.

2. Early- and Late-Stage Convolution Model

2.1 *Model Definition*

The standard convolution model assumes a preclinical (asymptomatic) and a clinical (symptomatic) disease state. An augmented model includes two preclinical and two clinical states as illustrated in Figure 1. The model depicted in

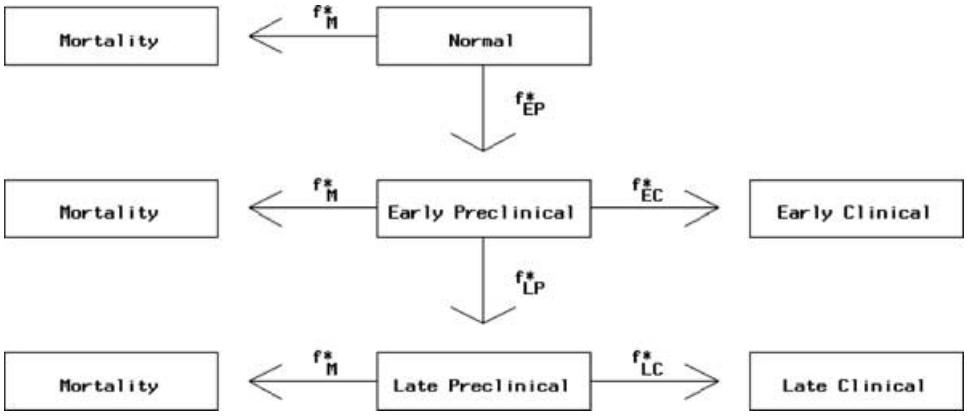

Figure 1. Schematic of early- and late-stage convolution model, with competing mortality from other causes.

Figure 1 also includes mortality from other causes, i.e., causes other than the disease of interest. All subjects are born into the normal state. In this state, subjects have a hazard $f^*_{\rm EP}(s)$ of entering the early preclinical state and a hazard $f^*_{\rm M}(s)$ of dying from other causes. Subjects entering the early preclinical state at age s have competing hazards $f^*_{\rm LP}(t-s)$ of entering the late preclinical state at age t, $f^*_{\rm EC}(t-s)$ of entering the early clinical state at age t, and $f^*_{\rm M}(t)$ of dying from other causes. A subject entering the late preclinical state at age x has a hazard of $f^*_{\rm LC}(t-x)$ of entering the late clinical state at age t and a hazard $f^*_{\rm M}(t)$ of dying from other causes. We assume that mortality from other causes is independent of the incidence (preclinical or clinical) of the disease of interest.

From basic probability theory, if a population entering a given state (state 0) at time 0 is exposed to independent hazards $f^*_1(t),\ldots,f^*_k(t)$ of transitioning into states 1 to k, then the proportion still in state 0 at time t will be $A(t)=\Pi_{i=1,k}\exp(-\int_0^t f^*_i(s)\,ds)$, and the density of subjects entering state j at time t will be $f^*_j(t)A=f_j(t)\sum_{i\neq j}(1-F_j(t))$, where $1-F_j=\exp(-\int_0^t f^*_j(s)\,ds)$ and $f_j=f^*_j(1-F_j)$ (i.e., f_j is the probability density and F_j the cumulative distribution function corresponding to the hazard f^*_j). Then it follows that the density of incident early- and late-stage disease are given by the following integral equations:

$$I_0(t)=\int_0^t f_{\rm EP}(s)F^C_{\rm M}(s)f_{\rm EC}(t-s)F^C_{\rm LP}(t-s)\times\left\{F^C_{\rm M}(t)\big/F^C_{\rm M}(s)\right\}ds \tag{1}$$

$$=F^C_{\rm M}(t)\int_0^t f_{\rm EP}(s)f_{\rm EC}(t-s)F^C_{\rm LP}(t-s)\,ds, \tag{2}$$

$$I_1(t)=\int_0^t f_{\rm EP}(s)F^C_{\rm M}(s)\left[\int_s^t f_{\rm LP}(x-s)F^C_{\rm EC}(x-s)\times\left\{F^C_{\rm M}(x)\big/F^C_{\rm M}(s)\right\}f_{\rm LC}(t-x)\,dx\right]\times\left\{F^C_{\rm M}(t)\big/F^C_{\rm M}(x)\right\}ds \tag{3}$$

$$=F^C_{\rm M}(t)\int_0^t f_{\rm EP}(s)\int_s^t f_{\rm LP}(x-s)\times F^C_{\rm EC}(x-s)f_{\rm LC}(t-x)\,dx\,ds, \tag{4}$$

where $F^C_X=1-F_X$.

Letting $I(t)=I_0(t)+I_1(t)$ be the total incidence density of disease, it is easily seen by summing equations (2) and (4) above that

$$I(t)=F^C_{\rm M}(t)\int_0^t f_{\rm EP}(s)f_{\rm C}(t-s)\,ds, \tag{5}$$

where

$$f_{\rm C}(t-s)=f_{\rm EC}(t-s)F^C_{\rm LP}(t-s)+\int_s^t f_{\rm LP}(x-s)F^C_{\rm EC}(x-s)f_{\rm LC}(t-x)\,dx. \tag{6}$$

2.2 *Probability of Disease in a Screened Population*

The above model can be utilized to compute the probability of finding disease, in a clinical or preclinical state, in a population undergoing screening. A subject undergoing screening may be preclinical (asymptomatic), and have disease identified on the basis of a positive screening test and resultant diagnostic workup, or may have disease identified clinically on the basis of symptoms. The former are denoted as prevalent and the latter as incident cases. Generally, the sensitivity of a screening test is defined as the probability that a subject with preclinical disease will have a "positive" screening test. Here, since there are two stages of preclinical disease, we postulate that there may be two different sensitivities, say B_0 and B_1 for early- and late-stage preclinical disease, respectively.

Using the natural history model and the screening test sensitivities, we can calculate the probability that a subject is identified as a prevalent case at a given screening visit and the probability that a subject is observed as an incident case in between screening visits or after the last scheduled screen. Among the population undergoing the jth screen, we denote by P^k_j the probability of being a prevalent case in stage k ($0=$ early, $1=$ late) at the jth screen and by $J^k_j(t)$ the density of stage k incident cases after the jth screen (and before any other screens). The screening times (ages) are denoted

as $t_1, \ldots, t_K$, with $t_0 = 0$ and t_{K+1} being an age such that $F_{\mathrm{M}}(t_{K+1}) = 1$. It can be shown then that P and J are given as follows:

$$P_j^0 = \sum_{m=0}^{j-1} \int_{t_m}^{t_{m+1}} f_{\mathrm{EP}}(s) F_{\mathrm{LP}}^C(t_j - s) \times F_{\mathrm{EC}}^C(t_j - s)(1 - B_0)^{j-m-1} B_0 \, ds, \quad (7)$$

$$P_j^1 = \sum_{m=0}^{j-1} \int_{t_m}^{t_{m+1}} f_{\mathrm{EP}}(s) \sum_{n=m}^{j-1} \int_{\max(s,t_n)}^{t_{n+1}} \times f_{\mathrm{LP}}(x - s) F_{\mathrm{EC}}^C(x - s) F_{\mathrm{LC}}^C(t_j - x) \, dx \times (1 - B_0)^{n-m}(1 - B_1)^{j-n-1} B_1 \, ds, \quad (8)$$

$$J_j^0(t) = \sum_{m=0}^{j} \int_{t_m}^{t_{m+1}} f_{\mathrm{EP}}(s) F_{\mathrm{LP}}^C(t - s) \times f_{\mathrm{EC}}(t - s)(1 - B_0)^{j-m} \, ds, \quad t \in (t_j, t_{j+1}), \quad (9)$$

$$J_j^1(t) = \sum_{m=0}^{j} \int_{t_m}^{t_{m+1}} f_{\mathrm{EP}}(s) \sum_{n=m}^{j} \int_{\max(s,t_n)}^{t_{n+1}} \times f_{\mathrm{LP}}(x - s) F_{\mathrm{EC}}^C(x - s) f_{\mathrm{LC}}(t - x) \, dx \times (1 - B_1)^{j-n}(1 - B_0)^{n-m} \, ds, \quad t \in (t_j, t_{j+1}). \quad (10)$$

To derive the expression for P_j^1, note that the subject has to have entered the early preclinical state and then entered the late preclinical state before the time of the screen, and cannot have exited the late preclinical state before the screen; in addition, if the subject was in a preclinical state at the times of prior screens *1* to $j - 1$, the screening test had to falsely call them negative (which happens with probability $1 - B_0$ or $1 - B_1$ depending on whether the preclinical state is early or late), and they had to have a positive test at the *j*th screen.

Fitting this model to screening trial data requires the number of subjects screened and the number of prevalent early- and late-stage cases at each screen. In addition, for each screen, one needs the number of person years of follow-up for different time periods following that screen (and before any other screens), with the corresponding number of early- and late-stage incident cases. Table 1 gives an idea of the type of data needed for a study with K screening exams; the intervals after each screen have been divided up here into L periods (e.g., 0–1, 1–2, and 2–6 years for $L = 3$). Optimally, these data would be available by (say) five-year age group. It is assumed that the time interval between the *j*th and $(j + 1)$th screen is relatively constant across the study. The expected number of incident stage k cases occurring during the *i*th time interval after the *j*th screen can be approximated as a Poisson random variable with mean $J_j^k(\bar{t}_{ij})Y_{ij}$, where Y_{ij} is the number of person years at risk during the interval as displayed in Table 1 and $\bar{t}_{ij}$ is the midpoint of the interval.

Technically, the above expressions for P (prevalent cases) and J (incident cases) consider all living subjects to be at risk, whereas in most studies only those subjects with no history of disease will be included in the cohort. For a typical relatively rare disease however, the difference will be negligible. Also, we exclude competing mortality from the expressions for P and J because we assume subjects are alive at the time of the screen and because deaths are subtracted from the number of person years at risk when calculating expected incident cases. However, when using the above formulas to predict the absolute number of prevalent and incident cases in a cohort followed from a given age, competing mortality should be added to the expressions for P and J as appropriate.

Table 1
Data from screening study needed for model fitting

	Screening exam number	1	...	K
No. Screened		N_1		N_K
Prevalent cases	Early stage	P_{01}	...	P_{0K}
	Late stage	P_{11}	...	P_{1K}
Period 1	Person years	Y_{11}		Y_{1K}
	Incident cases—early stage	I_{011}		I_{01K}
	Incident cases—late stage	I_{111}		I_{11K}
...				
Period L	Person years	Y_{L1}		Y_{LK}
	Incident cases—early stage	I_{0L1}		I_{0LK}
	Incident cases—late stage	I_{1L1}		I_{1LK}

Specific functions (e.g., exponential, gamma, Weibull) can be chosen for f_{LC}, f_{EC}, and f_{LP} to make the model parametric. As demonstrated by Pinsky (2001) for the standard two-state convolution model, by assuming that the incidence density in the absence of screening, $I(t)$, is known and prespecified, the preclinical incidence function effectively becomes a function of the clinical incidence density through deconvolution. In the current augmented model this also holds; specifically, here f_{EP} becomes a function of $I = I_0 + I_1$ and the clinical incidence density f_{C} (where f_{C} is defined by [6]) through deconvolution of equation (5). Thus if I is assumed known, the model parameters needed to be fit are effectively B_0, B_1, and the parameters of f_{LC}, f_{EC}, and f_{LP}; the expression for f_{EP} in the likelihood is determined from I and f_{C}, where the latter is itself a function of f_{LC}, f_{EC}, and f_{LP}. Alternatively, one could fit f_{EP} directly by parameterizing it, or its hazard, as a polynomial or other function. Whether to fit f_{EP} directly or to utilize a known incidence rate (if available) and deconvolution is in some ways a tradeoff between bias and variance, with the former method minimizing bias and the latter minimizing variance (Pinsky, 2001).

Finally, we briefly discuss identifiability of the model parameters. Let f_{EC}, f_{LP}, and f_{LC} (and f_{EP} if it is being fit) be parameterized by θ_{EC}, θ_{LP}, etc. We assume that θ_{EC}, etc., is an identifiable parameter for f_{EC} in the sense that $f_{\mathrm{EC}}(\theta_1) = f_{\mathrm{EC}}(\theta_2)$ only if $\theta_1 = \theta_2$. We also assume that the observable data are in the form of Table 1, with a total number of incident and prevalent case cells of n. Then the expected value of the observable data will be a vector, say W, derived from the expressions for P_j^0, P_j^1, $J_j^0(t)$, and $J_j^1(t)$. In general the Jacobian of W as a function of the parameter vector θ (which is of dimension m and includes B_0, B_1, θ_{EC}, θ_{LP}, θ_{LC}, and θ_{EP}, if it is being fit) will be of full rank; thus, as long as $m \leq n$, the model parameters will generally be identifiable (Bamber and van Santen, 2000).

3. Calculating Sojourn Time, Lead Time, and Overdiagnosis

3.1 *Sojourn Time*

The sojourn time of a disease is the time that the subjects remain in the preclinical disease state. Three sojourn time distributions are of interest in the current model. First, there is the time in the early preclinical state for subjects who eventually present clinically with early-stage (stage 0) disease; this sojourn time distribution is denoted by S^0. It is straightforward to show that S^0 is given by

$$S^0(z) = \frac{\displaystyle\int_0^{t^F} f_{\mathrm{EP}}(s) f_{\mathrm{EC}}(z) F^C_{\mathrm{LP}}(z) F^C_{\mathrm{M}}(s+z)\, ds}{\displaystyle\int_0^{t^F} I_0(t)\, dt}, \qquad (11)$$

where t^F is an age such that $F_{\mathrm{M}}(t^F) = 1$. For cases who present clinically with late-stage disease, we are interested in the joint distribution, $S^1(x, z)$, of the sojourn times in the early (x) and late (z) preclinical states. $S^1(x, z)$ is then given as follows:

$$S^1(x,z) = \frac{\displaystyle\int_0^{t^F} f_{\mathrm{EP}}(s) f_{\mathrm{LP}}(x) F^C_{\mathrm{EC}}(x) f_{\mathrm{LC}}(z) F^C_{\mathrm{M}}(z+x+s)\, ds}{\displaystyle\int_0^{t^F} I_1(t)}. \qquad (12)$$

The above sojourn time distributions characterize preclinical cases who would present clinically in the absence of intervention. The complement of these are the preclinical cases who never become clinical due to competing mortality. With no screening, the proportion of preclinical cases who never become clinical, $1 - P_{\mathrm{Clin}}$, is given by

$$1 - P_{\mathrm{Clin}} = 1 - \frac{\displaystyle\int_0^{t^F} I(t)}{\displaystyle\int_0^{t^F} f_{\mathrm{EP}}(t) F^C_{\mathrm{M}}(t)\, dt}. \qquad (13)$$

Note that the denominator in the above expression is the proportion of a cohort of subjects ever developing preclinical disease.

3.2 *Lead Time and Overdiagnosis*

The current model can also be used to compute the probabilities of various outcomes for screen-detected cases, assuming they received no screen-related intervention. Note that such outcomes are usually not directly observable, because screen-detected cases usually do receive some intervention. Among subjects screen diagnosed in the early and/or late stage, we can calculate the probability that the subject would have clinically presented in the early and/or late stage, or never clinically presented, in the absence of intervention. The term overdiagnosis measures the extent to which screen-detected cases would never have presented clinically. In the current model formulation, the extent of overdiagnosis is a function of the competing hazards of clinical presentation on the one hand versus other cause mortality on the other; decreasing the former relative to the latter will increase the overdiagnosis rate. Among the fraction of screen-detected cases who would be expected to have presented clinically, the "lead time" is the time interval by which diagnosis is speeded up, i.e., the time between screen-related diagnosis and when one would have presented clinically.

We let $W_j^{00}(x)$ denote the lead-time density for cases screen detected at the jth screen in stage 0 (early stage) who would have presented in stage 0 (in the absence of intervention). Specifically, $W_j^{00}(x)$ is the probability of being screen detected at the jth screen, at age t_j, and having been destined to clinically present (in the absence of intervention) in stage 0 at age $t_j + x$. Similary, $W_j^{11}(y)$ is the lead-time density for cases screen detected in stage 1 (late stage) who would have presented clinically in stage 1. Finally, $W_j^{01}(x,\ y)$ is the joint density of lead times x and y in the early and late stage, respectively, of subjects screen detected in the early stage who would have presented clinically in the late stage. The (improper) lead-time densities W_j^{ik} are then calculated as follows:

$$\begin{aligned}
W_j^{00}(x) &= \sum_{m=0}^{j-1} \int_{t_m}^{t_{m+1}} f_{\mathrm{EP}}(s) f_{\mathrm{EC}}(x + t_j - s) \\
&\quad \times F^c_{\mathrm{LP}}(x + t_j - s) F^c_{\mathrm{M}}(t_j + x)(1 - B_0)^{j-m-1} B_0\, ds, \\
W_j^{01}(x,y) &= \sum_{m=0}^{j-1} \int_{t_m}^{t_{m+1}} f_{\mathrm{EP}}(s) f_{\mathrm{LP}}(t_j + x - s) F^c_{\mathrm{EC}}(t_j + x - s) \\
&\quad \times f_{\mathrm{LC}}(y) F^c_{\mathrm{M}}(t_j + x + y)(1 - B_0)^{j-m-1} B_0\, ds, \\
W_j^{11}(y) &= \sum_{m=0}^{j-1} \int_{t_m}^{t_{m+1}} f_{\mathrm{EP}}(s) \sum_{n=m}^{j-1} \left\{ \int_{\max(s,t_n)}^{t_{n+1}} f_{\mathrm{LP}}(x - s) \right. \\
&\quad \times F^c_{\mathrm{EC}}(x - s) f_{\mathrm{LC}}(t_j + y - x) \\
&\quad \left. F^c_{\mathrm{M}}(t_j + y)\, dx \right\} (1 - B_1)^{j-n-1} (1 - B_0)^{n-m} B_1\, ds.
\end{aligned}$$

Note the t_i are again the screening ages, with $t_0 = 0$. The (improper) lead-time density $W^{im}(y)$ over all screens is computed then by summing up $W_j^{im}(y)$ over all screens $j = 1, K$.

We define C_j^{im} as the proportion of those screen diagnosed at the jth screen in stage i who would have eventually been clinically diagnosed (in the absence of intervention) in stage m. Here, i and m can be either 0 (early), 1 (late), or x (early or late). Various C_j^{im} of interest are given below:

$$\begin{aligned}
C_j^{01} &= \frac{\displaystyle\int_0^{t^F}\int_0^{t^F} W_j^{01}(x,y)\, dx\, dy}{F^c_{\mathrm{M}}(t_j) P_j^0}, \\
C_j^{11} &= \frac{\displaystyle\int_0^{t^F} W_j^{11}(y)\, dy}{F^c_{\mathrm{M}}(t_j) P_j^1}, \\
C_j^{x1} &= \frac{P_j^0 C_j^{01} + P_j^1 C_j^{11}}{P_j^0 + P_j^1}, \\
C_j^{xx} &= \frac{\displaystyle\int_0^{t^F} W_j^{00}(y)\, dy + F^c_{\mathrm{M}}(t_j)\left(P_j^0 + P_j^1\right) C_j^{x1}}{F^c_{\mathrm{M}}(t_j)\left(P_j^0 + P_j^1\right)}.
\end{aligned}$$

Note that the denominator for C^{01} and C^{11} is the total number of screen-detected cases in the appropriate stage while the numerator is the number of such cases who would have eventually clinically presented. To get C^{im} over all screens, one simply takes the weighted average of the C_j^{im}, with the weights being the number of screen-identified cases in the appropriate stage at the jth screen (i.e., the denominator of the above expressions).

4. Example—Fitting Model to Screening Trial Data

4.1 *Mayo Lung Trial Description*

We fit the model using maximum likelihood to data from the Mayo Lung Cancer Screening Trial (Fontana et al., 1984). Men aged 45 and over who smoked at least one pack of cigarettes daily (either at the time of enrollment or during the previous year) were eligible for the trial. All trial subjects received an initial "prevalence" screen consisting of a chest X-ray and sputum cytology. After the initial screen, study subjects not diagnosed with cancer, and having adequate life expectancy and respiratory reserve, were randomized to a screened or control arm. Men in the screened arm received chest X-ray and sputum cytology tests every four months for a six-year period, with an additional two years on average of follow-up after the last scheduled screen. Control arm subjects received usual Mayo clinic care, which included a recommended annual chest X-ray and cytology. Because we do not have data on the number and timing of any screens in the control group, we did not attempt to model the control group experience. We included all the data on subjects undergoing the prevalence screen and all the data for the subsequent experience of the screened arm subjects; nonscreened arm subjects who had the prevalence screen were censored after this initial screen.

We define early-stage lung cancer here as American Joint Committee on Cancer (AJCC) stage I–II and late stage as AJCC stage III–IV. Study researchers also classified each case as being screen detected (prevalent) or symptom identified (incident). In addition, some cases were classified as being diagnosed incidentally from nonstudy chest X-rays obtained for other clinical reasons. For modeling purposes, incidental cases were considered screen detected. However, since the nature of the incidental exam may have differed from the standard study screening exam and since it is unknown how many other subjects received such an exam, we used "nuisance" sensitivity parameters to model the risk of incidental cases. These parameters have little intrinsic meaning since they are a combination of the frequency of incidental exams and the incidental exam sensitivity.

To parameterize the functions f_{EC}, f_{LC}, and f_{LP}, we assumed each had the general form of a Weibull distribution. If the Weibull shape parameter α for any of the functions was not significantly different from 1, then the simpler exponential distribution was used for that function in the final model. As mentioned above, if the incidence $I(t) = I_0(t) + I_1(t)$ in the absence of screening is assumed known, then f_{EP} becomes a function of f_{EC}, f_{LC}, f_{LP}, and I through deconvolution. For a cancer screening trial involving the general population, population registries for a time or place with no screening may provide a suitable data source for I. However, this population was limited to heavy smokers, who have a greatly different lung cancer risk than the general population. Therefore, we did not assume that $I(t)$ was known, but instead fit f_{EP} directly by assuming that its hazard as a function of age was a cubic polynomial (with no constant term).

Table 2
Observed and predicted case breakdown, Mayo Lung Trial

		Observed		Predicted	
		Early	Late	Early	Late
Prevalent cases	1st screen	45	46	42	39
Prevalent cases	Later screens	65	25	68	30
Incident cases	0–2 years	5	30	3	23
Incident cases	2+ years	2	36	3	36
Incidental cases		26	17	26	17

Note: All 10,933 study subjects received first screen; after this screen noncases were randomized to screened or control arm. Later screen prevalent, incident, and incidental cases are restricted here to screened arm subjects ($n = 4618$). Incident cases are broken down according to whether diagnosis was 0–2 years or 2+ years after the last screen.

We prepared the original Mayo data in the form of Table 1, with separate tables for each five-year age group. We used the following three periods for incident cases after the final screen: 0–2, 2–4, and 4–10 years.

4.2 *Results*

A total of 10,933 subjects received the initial prevalence screen and 4618 subjects were randomized to the screened arm. The mean age of study participants was 55 and the mean number of pack years of smoking was 49. Table 2 gives a breakdown of the observed prevalent and incident cases by stage (early or late) in the study (Fontana et al., 1986). Only 10% of the incident cases were early stage (AJCC stage I–II), while 49% of the cases identified at the initial screen and 72% of the cases identified at later screens were early stage.

The maximum likelihood parameter estimates for the model fit to the Mayo data are displayed in Table 3. The p-value for the null hypothesis that the Weibull parameter α was equal to one was 0.0001 for f_{LP}, 0.10 for f_{EC}, and 0.14 for f_{LC}; thus the final model included a full Weibull distribution

Table 3
ML parameter estimates of model fit to Mayo Lung Trial data

Function	Parameter	ML estimate	95% CI
f_{EC}	λ_{EC} (year^{-1})	0.068	(0.03, 0.13)
f_{LC}	λ_{LC} (year^{-1})	0.85	(0.62, 1.1)
f_{LP}	λ_{LP} (year^{-1})	1.03	(0.80, 1.2)
f_{LP}	α	0.57	(0.40–0.77)
f_{EP}	c_1	−0.013	
	c_2	0.023	
	c_3	0.0043	
Early-stage sensitivity	B_0	1.0	(0.74, 1.0)
Late-stage sensitivity	B_1	1.0	(0.80, 1.0)

Note: f_{EC} and f_{LC} are exponential distributions with parameters λ_{EC} and λ_{LC}. f_{LP} is a Weibull distribution with parameters λ_{LP} and α. The hazard associated with f_{EP} was modeled as $\mathrm{Max}(0,\ c_1T + c_2T^2 + c_3T^3)$, where T is the age divided by 80.

for f_{LP} and exponential distributions for f_{EC} and f_{LC}. The (constant) hazard for transitioning into the early clinical stage, 0.068 (year^{-1}), was much lower than the (constant) hazard for transitioning into the late clinical state, 0.85. The hazard for transitioning into the late preclinical state was nonconstant and decreasing over time; the average hazard was 1.04 in the first year and 0.51 in the second year.

The sensitivity estimate for both late- and early-stage disease was 1.0. Note the sensitivity parameter here is for the combined screen of chest X-ray plus sputum cytology. Of screen-detected cases, chest X-ray was positive for 73% of early-stage and 93% of late-stage cases, so a sensitivity estimate for chest X-ray alone would be 0.73 and 0.93 for early- and late-stage cancer, respectively. The corresponding sensitivity estimates for cytology alone were 0.33 and 0.29 for early- and late-stage disease, respectively.

Table 2 displays the case breakdown predicted by the model, as well as the observed breakdown. The model fits the overall pattern quite well. Table 4 displays estimated characteristics of a hypothetical cohort of heavy-smoking men who undergo yearly screening with the Mayo regimen (i.e., chest X-ray and cytology) from age 50 to 75. The numbers were computed using the above-fitted parameters and another cause mortality hazard derived from the Mayo trial cohort. The average sojourn time for subjects clinically presenting with early-stage disease was 1.8 years. For subjects clinically presenting with late-stage disease, the average sojourn times in the early and late preclinical states were 0.85 and 1.05 years. Mean lead times are slightly longer. The proportion of cases screen detected in early stage that would have presented clinically in late stage (C^{00}) was 0.83 and the proportion of all screen-detected cases that would have presented clinically in any stage (C^{xx}) was 0.94.

Table 4
Estimated characteristics of cohort undergoing annual lung screening with Mayo screening regimen

Measure	Quantity	Value
Mean sojourn time (year)	$\bar{S}^0$	1.8
	$\bar{S}^1$ [early]	0.85
	$\bar{S}^1$ [late]	1.05
Mean lead time (year)	$\bar{W}^{00}$	2.1
	$\bar{W}^{01}$ [early]	1.4
	$\bar{W}^{01}$ [late]	1.1
	$\bar{W}^{11}$	1.1
Overdiagnosis	C^{01}	0.83
	C^{11}	0.95
	C^{xx}	0.94
% Ever in preclinical state		14.0
% w/diagnosed cancer		13.7
% w/screen-detected cancer		8.8
% w/diagnosed late-stage cancer		8.2

S^i is sojourn time in preclinical state for subjects clinically presenting in stage i. W^{ij} is lead time of subjects screen detected in stage i who would have clinically presented in stage j. C^{ij} is proportion of subjects in stage i who would have presented in stage j in the absence of screening. Code for stage: 0 = early, 1 = late, x = early or late. Cohort assumed to undergo annual screening from age 50 to 75.

In this example the lead times and the proportions C^{ij} are relatively independent of the screening frequency. In contrast, the proportion of the cohort diagnosed with late-stage disease is quite sensitive to the screening frequency. If screening frequency were increased to twice annually, the proportion of the cohort ever diagnosed with late-stage disease rate would drop from 8.2% to 7.2%, while with no screening the proportion would increase to 12.1%.

The values for C^{ij} presented in Table 4 indicate a relatively low level of overdiagnosis for a population undergoing annual screening, in the sense that a high proportion of those with screen-detected disease would have been expected to ever present with clinical disease if there had been no screening. This measure of overdiagnosis is related to, but not exactly equivalent to, the overdiagnosis rate in a randomized trial, which can be defined as the percent increase in the disease rate in the screened group relative to the control group. The Mayo Lung Trial reported an overdiagnosis rate of 28% (Fontana et al., 1986). Under the conditions of the Mayo Trial, i.e., an initial prevalence screen, the given age distribution, six years of screening, and an additional two years on average of postscreening follow-up, the fitted model (which did not utilize control group data for fitting) predicts 144 incident cases in the control group, as compared to 169 screen-detected plus incident cases in the screened group. Assuming the same number of incidental cases in each group, the predicted overdiagnosis rate would be 13%; ignoring incidental cases would give a predicted overdiagnosis rate of 17%.

To examine the robustness of the model results we utilized an alternate form of the early preclinical incidence function f_{EP}. In addition to using the cubic polynomial hazard, we also fit the overall model using an exponential hazard for f_{EP}, i.e., letting the hazard equal $\exp(c_0 + c1t + c2t^2)$, where t is the age. The resulting maximum likelihood estimates of the other parameters in the model (e.g., the exponential parameter for f_{EC}, etc.) were essentially unchanged by using the exponential hazard, and the model had a similar but slightly worse fit in terms of the log likelihood. However, the early preclinical incidence curves diverged in the two models after about age 70, with the exponential curve increasing at a greater rate than the polynomial curve. This age range is largely outside that seen in the Mayo Trial; however, this highlights the uncertainties involved in extrapolating beyond the age range of the data that were used to fit the model. The model with the exponential f_{EP} hazard gives somewhat greater lifetime clinical and preclinical disease probabilities than those displayed in Table 4, which were derived using the polynomial hazard. The polynomial hazard is probably more biologically plausible in terms of the rate at which cancer risk increases with age in the older age range.

5. Discussion

We have shown here how a natural history model that incorporates two preclinical, as well as two clinical, states can be fit using maximum likelihood to data from a screening study. We have also shown how the resulting parameter estimates can be used to extrapolate the findings of the screening study to related settings. With this model, one can consider the tradeoff between the frequency of screening and various outcomes, such as the diagnosis rate of early- and of late-stage disease.

Use of this model for such an analysis allows for parameters which can be estimated directly from data, with corresponding confidence intervals. The current model, however, does not consider disease survival, which is an important component of cost-effectiveness. An additional component then would involve modeling disease survival as a function of stage and method of diagnosis (screen detected or incident).

While the current model is more complex than the simple two-state convolution model that has been widely used in the past, it is less complex than other natural history models which incorporate a large number of preclinical disease states, such as the models developed by the MISCAN group (van der Maas et al., 1989; Draisma et al., 2003). While these more complex models allow for a more detailed description of possible preclinical states and related transitions, their complexity comes at a cost of the model parameters not being identifiable from available screening trial data. For example, the MISCAN prostate cancer model has nine preclinical states corresponding to three different stages of cancer crossed with three different Gleason grades (Draisma et al., 2003). Church (2003) argues that while this model is useful, the large number of parameters compared to the number of data cells being fit makes it likely that the model fit is underdetermined, i.e., that the model parameters are not identifiable.

When this model is fit to screening trial data, the estimated parameters should be interpreted in the appropriate context. For example, a positive screening test is normally followed by a diagnostic work-up where a standard definitive procedure such as biopsy is used for disease diagnosis. The estimated preclinical incidence function, f_{EP}, and other relevant parameters must be interpreted within this context. Thus, from the point of view of the fitted model parameters, the onset of preclinical disease is defined as the first point where the tumor (say) could generally be diagnosed with a biopsy or other standard diagnostic technique, not the first biological emergence of the tumor. If a new technology were to come along that could diagnose smaller tumors, then a different preclinical incidence function could be expected.

The model as fit to the Mayo Lung Trial data captured the essential pattern of disease presentation, including the fact that the proportion of cases which were diagnosed in the late stage was greatest for the incident cases, next greatest for the prevalent cases identified on the initial screen, and least for the prevalent cases identified after the first screen. The parameter estimates obtained are generally consistent with the accepted picture of lung cancer as a rapidly progressing disease. This rapid progression dictates that a large majority of preclinical cases are predicted to progress to clinical cases in the absence of intervention, and usually to late-stage disease. On the other side of the spectrum would be a disease such as prostate cancer, which has been estimated to have a long lead time (Auvin et al., 2002). For prostate cancer, one would expect the model to predict that relatively few early stage (AJCC stage I–II) screen-detected cancers would have presented clinically in late stage in the absence of screening.

Biologically, one would generally expect that the hazard for symptoms (i.e., for presenting clinically) would be smaller for early preclinical than for late preclinical disease and that the sensitivity for early-stage disease would be less than that for late-stage disease. The model, however, is not constrained in any way regarding these parameters. In the Mayo example, the former of these expectations was clearly observed. With respect to the latter, while the ML estimate of early-stage sensitivity was not lower than, but equivalent to, the ML estimate of late-stage sensitivity, the lower 95% CI was lower for early- than for late-stage sensitivity (each had an upper 95% CI of 1.0). This provides additional evidence for the plausibility of the model in this example. In the Mayo example we also found that using a Weibull as opposed to an exponential function to model the transition from early to late preclinical state significantly improved the model fit. The Weibull hazard of transitioning to the late preclinical state was decreasing over time, which is consistent with the idea that there may be some less aggressive tumors which do not rapidly progress to late stage, as well as the majority of aggressive tumors that do.

Résumé

Le modèle standard de convolution d'une histoire naturelle d'une maladie pose en principe l'existence de deux états : l'un sans symptôme (pré-clinique) et l'autre avec symptôme (clinique). Un modèle augmenté inclut, dans les deux états pré-clinique et clinique, une phase précoce et une phase tardive de la maladie. Dans le cas d'un cancer, la phase précoce correspondrait généralement aux périodes limitées à un organe avant qu'il n'y évidence de la diffusion. Nous calculons le nombre attendu de symptômes détectés (pré-clinique) et celui de cas cliniques dans les phases précoces et tardives dans un programme donné de détection et nous montrons comment le modèle peut être ajusté aux données d'un essai de détection par maximum de vraisemblance. Nous développons aussi des expressions pour le temps de séjour, le temps de début et de sur-diagnostic dans le contexte du modèle, où chacun des concepts précédents incorpore la phase de la maladie. Comme exemple, nous ajustons le modèle aux données d'un essai de détection du cancer du poumon.

References

Auvin, A., Maattanen, L., Stenman, U. H., Tammela, T., Rannikko, S., Aro, J., Juusela, H., and Hakama, H. (2002). Lead-time in prostate cancer screening (Finland). *Cancer Causes and Control* **13,** 279–285.

Bamber, D. and van Santen, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology* **44,** 20–40.

Chen, H. H., Duffy, S. W., and Tabar, L. (1996). A Markov chain method to estimate the tumor progression rate from preclinical to clinical phase, sensitivity and positive predictive value for mammography in breast cancer screening. *The Statistician* **45,** 307–317.

Chen, H. H., Kuo, H. S., Yen, M. F., Lai, M. S., Tabar, L., and Duffy, S. W. (2000). Estimation of sojourn time in chronic disease screening without data on interval cases. *Biometrics* **56,** 167–172.

Church, T. (2003). Simulated screening for prostate cancer: The useful model. *Journal of the National Cancer Institute* **95,** 838–839.

Day, N. E. and Walter, S. D. (1984). Simplified models of screening for chronic disease: Estimation procedures from mass screening programmes. *Biometrics* **40,** 1–14.

Draisma, G., Boer, R., Otto, S. J., van der Cruijsen, I. W., Damhuis, R. A. M., Schroder, F. H., and de Koning, H. J. (2003). Lead times and overdetection due to prostate-specific antigen screening: Estimates from the European Randomized Study of Screening for Prostate Cancer. *Journal of the National Cancer Institute* **95,** 868–877.

Fontana, R. S., Sanderson, D. R., Taylor, W. F., Woolner, L. B., Miller, E., Muhm, J. R., and Uhlenhopp, M. (1984). Early lung cancer detection: Results of the initial (prevalence) radiologic and cytologic screening in the Mayo Clinic Study. *American Review of Respiratory Disease* **130,** 561–565.

Fontana, R. S., Sanderson, D. R., Woolner, L. B., Taylor, W. F., Miller, E., and Muhm, J. R. (1986). Lung cancer screening: The Mayo program. *Journal of Occupational Medicine* **28,** 746–750.

Paci, E. and Duffy, S. W. (1991). Modeling the analysis of breast cancer screening programmes: Sensitivity, lead time and predictive value in the Florence district programme (1975–1986). *International Journal of Epidemiology* **20,** 852–858.

Pinsky, P. (2001). Estimation and prediction for cancer screening models using deconvolution and smoothing. *Biometrics* **57,** 389–395.

van der Maas, P. J., de Koning, H. J., van Ineveld, B. M., van Oortmartssen, G. J., Habbema, J. D., Lubbe, K. T., Geerts, A. T., Collette, H. J., Verbeek, A. L., and Hendriks, J. H. (1989). The cost-effectiveness of breast cancer screening. *International Journal of Cancer* **43,** 1055–1060.

Received April 2003. *Revised August* 2003.
Accepted October 2003.