BIOMETRICS 57, 389–395
June 2001

Estimation and Prediction for Cancer Screening Models Using Deconvolution and Smoothing

Paul F. Pinsky

Division of Cancer Prevention, National Cancer Institute,
EPN 330, 6130 Executive Boulevard, Bethesda, Maryland 20892, U.S.A.
email: pp4f@nih.gov

SUMMARY. The model that specifies that cancer incidence, I, is the convolution of the preclinical incidence, g, and the density of time in the preclinical phase, f, has frequently been utilized to model data from cancer screening trials and to estimate such quantities as sojourn time, lead time, and sensitivity. When this model is fit to the above data, the parameters of f as well as the parameter(s) governing screening sensitivity must be estimated. Previously, g was either assumed to be equal to clinical incidence or assumed to be a constant or exponential function that also had to be estimated. Here we assume that the underlying incidence, I, in the study population (in the absence of screening) is known. With I known, g then becomes a function of f, which can be solved for using (numerical) deconvolution, thus eliminating the need to estimate g or make assumptions about it. Since numerical deconvolution procedures may be highly unstable, however, we incorporate a smoothing procedure that produces a realistic g function while still closely reproducing the original incidence function I upon convolution with f. We have also added the concept of competing mortality to the convolution model. This, along with the realistic preclinical incidence function described above, results in more accurate estimates of sojourn time and lead time and allows for estimation of quantities related to overdiagnosis, which we define here.

KEY WORDS: Cancer screening; Competing mortality; Convolution model; Deconvolution; Lead time; Overdiagnosis; Sojourn time.

1. Introduction

Screening programs for some types of cancer (e.g., breast, cervical) are currently an important part of the arsenal for secondary cancer prevention. Screening trials provide critical information on the assessment of screening modalities; however, they are limited in that they usually assess only a single screening regimen with a limited number of screens performed per subject. Mathematical models have frequently been fit to screening trial data to estimate various quantities such as lead time, sojourn time, and sensitivity, which in general cannot be directly measured; these quantities are helpful in extrapolating trial results to other contexts. Screening trial data is often of the form displayed data in Table 1, where the number of screen-detected (i.e., preclinical) and clinically detected cases and the corresponding number of person years (or screens) are shown for each of three age groups for a trial of fecal occult blood screening for colon cancer (Launoy et al., 1997).

Many researchers have utilized the standard convolution model of the form $I(t) = \int_0^t g(s)f(t-s)ds$ to analyze screening trial data, where $I(t)$ is the clinical incidence, g is the preclinical incidence, and f is the sojourn time density. Typically, simplifying assumptions about f and, especially, g have been made to simplify calculations or the fitting of the model. Day and Walter (1984), e.g., do not fit g but fix it at a constant rate (for each age group) equivalent to the existing (age group-specific) clinical incidence rate; f, which is assumed exponential, was fit to the data along with a parameter denoting sensitivity. Others have fit an exponential g and f to screening data (e.g., Duffy et al., 1995; Chen, Duffy, and Tabar, 1996) or fit a step function g (Shen and Zelen, 1999).

In this article, we introduce two new concepts that can be utilized in conjunction with the standard convolution model for the purposes of estimating quantities related to screening. The first is to assume that the incidence rate in the absence of screening (i.e., $I(t)$ above) is known. As we will show in this article, this assumption allows one to utilize numerical deconvolution (and smoothing) techniques that effectively render g a function of f, thus precluding the need to parameterize and fit g. The second concept is to introduce competing mortality when estimating important quantities related to screening. Competing mortality implies that not all individuals with preclinical disease will eventually develop clinical disease (even in the absence of intervention). Including competing mortality allows one to quantify the important phenomenon of overdiagnosis, defined roughly as the tendency for screening to identify cases that would never have presented in the absence of screening; it also allows for more realistic estimations of such quantities as average lead time and sojourn time.

In Section 2, we review the basic convolution model of cancer incidence and describe how it is fit to data from cancer

Table 1
Fecal occult blood screening for colorectal cancer, Calvados, France

Age group		At screen	Years postscreen 1	2	3
45–54	Person years	25,041[c]	24,545	22,568	14,314
	Observed cases	12	4	3	2
	Expected cases[a]	17.3	5.0	6.1	4.8
	Expected cases[b]	10.6	5.7	6.8	5.2
55–64	Person years	24,933[c]	24,211	22,456	14,494
	Observed cases	35	16	7	8
	Expected cases[a]	46.3	13.0	15.7	12.3
	Expected cases[b]	41.8	14.2	16.5	12.7
65–74	Person years	19,836[c]	19,068	17,643	11,045
	Observed cases	84	24	19	12
	Expected cases[a]	70.5	18.4	20.5	14.9
	Expected cases[b]	83.7	19.7	20.9	14.7

[a] Expected cases based on model with $f(t-s \mid s)$ exponential with constant λ.
[b] Expected cases based on model with $f(t-s \mid s)$ exponential with time varying λ (see text).
[c] Person years here is number screened.

screening trials. Section 3 describes how, given f and I, deconvolution and smoothing can be used to determine the preclinical incidence function g; it also examines how misspecification of I affects parameter estimates. In Section 4, we introduce competing mortality to the model and show how lead time, sojourn time, length bias, and overdiagnosis are defined and calculated in the presence of competing mortality. Finally, in Section 5, we present an example of fitting this model to the data from Table 1 summarizing the results of a study of colon cancer screening.

2. Review of Basic Convolution Model Applied to Cancer Screening

The convolution model postulates that the clinical incidence, $I(t)$, of a given cancer in an unscreened cohort is determined by the following integral equation:

$$I(t) = \int_0^t g(s) f(t-s \mid s) ds, \tag{1}$$

where g is the preclinical incidence and f is the density of the length of the preclinical period, i.e., the period in which cases are potentially detectable through screening (Day and Walter, 1984). The model assumes that an individual in the preclinical state cannot return to the normal state without outside intervention. Note that $I(t)$ and $g(t)$ as defined are probability density functions, not hazards.

The likelihood for model (1) fit to data of the type shown in Table 1 is determined from expressions for the probability of being a screen-detected (i.e., prevalent) or clinically detected (i.e., incident) case at different time points for different age groups. These expressions have been derived before for special cases (e.g, Day and Walter, 1984), and, as the logical arguments are essentially the same, we will not repeat them. We generalize, however, on the results of other authors in the following three ways: First, we assume that the test sensitivity, $\beta = \beta(t-s)$, may be a function of the time $t-s$ since preclinical onset; second, we assume that the sojourn time density function, f, may depend on the age s of clinical onset; and third, we let g be a general function instead of a (piecewise) constant or an exponential density.

For the rest of this article, we suppose that screening occurs at ages $t_1, \ldots, t_k$ for the age group of interest. We denote by P_j the probability of being a prevalent case at the jth screen and by $J(t)$ the density of incident cases at time t. Then, similar to the results of Day and Walter (1984), these are given as follows:

$$\begin{aligned} P_j = &\sum_{m=0}^{j-2} \int_{t_m}^{t_{m+1}} g(s)(1 - F(t_j - s \mid s)) \\ &\times \left\{ \prod_{i=1}^{j-m-1} (1 - \beta(t_i - s)) \right\} \beta(t_j - s) ds \\ &+ \int_{t_{j-1}}^{t_j} g(s)(1 - F(t_j - s \mid s)) \beta(t_j - s) ds \end{aligned} \tag{2}$$

$$J(t) = \begin{cases} I(t), & t < t_1 \\ \displaystyle\sum_{m=0}^{j-1} \int_{t_m}^{t_{m+1}} g(s) f(t-s \mid s) \\ \qquad \times \left\{ \displaystyle\prod_{i=1}^{j-m} (1 - \beta(t_i - s)) \right\} ds \\ \qquad + \displaystyle\int_{t_j}^{t} g(s) f(t-s \mid s), & t \in (t_j, t_{j+1}), \\ & j = 1, k \end{cases} \tag{3}$$

where $t_0 = 0$, t_{k+1} equals the end of follow-up, and F is the cumulative distribution function (c.d.f.) of f. Since the case detection rate will generally be low, the likelihood can be approximated using a Poisson distribution. The expected number of incident cases in the ith time interval can be approximated as $J((t_i + t_{i+1})/2)N_i = J_i N_i$, where N_i is the number of person years at risk during the interval. The log

likelihood (LL) of $m_1, \ldots, m_k$ prevalent cases and $n_1, \ldots, n_k$ interval cases is then approximated by

$$LL = \sum_{j=1}^{k} m_j \log(P_j M_j) - P_j M_j + \sum_{i=1}^{k} n_i \log(J_i N_i) - J_i N_i, \tag{4}$$

where M_j is the number screened at t_j.

3. Utilizing Deconvolution and Smoothing to Specify g

The likelihood (4) is a function of f, g, and β. However, by assuming that $I(t)$ is known and prespecified, g effectively becomes a function of f through equation (1); thus, to maximize the likelihood, we only need to maximize over the parameters of f and β. Equation (1), with g unknown and I and f specified, is a Volterra equation of the first kind and, under general regularity conditions, it will have a unique continuous solution g^* (Linz, 1985). *A priori* knowledge of $I(t)$ thus provides a constraint on the relative values of f and g and, in theory at least, makes g a function of f. It is well known, however, that solutions of (1) may be highly nonsmooth and not necessarily nonnegative (O'Sullivan, 1986); therefore, many researchers employing deconvolution also incorporate a smoothing step. Suppose $S_h(\cdot)$ is a smoothing procedure with variable bandwidth h and let $g_h^* = S_h(g^*)$. Then, as h is increased from zero, the smoothness of g_h^* will generally increase but the error of $\int g_h^*(s) f(t-s \mid s)$ in reproducing $I(t)$ will also increase. Various authors have suggested ways to assess the trade-off between smoothness (or curvature) and error; Liao and Brookmeyer (1995), e.g., utilize a penalized likelihood approach.

Here we opt not to develop a rigorous algorithm for assessing this trade-off. Rather, as described in the Appendix, we have shown that, over a wide range of cancer incidence functions and f functions, a specific smoothing procedure (namely, a kernel smoothing routine with an Epanechnikov kernel and bandwidth anywhere from 1.5 to 2 years) is able to keep the error minimal while ensuring reasonable smoothness.

In reality, $I(t)$ will not be known exactly, and it is of interest to explore how misspecification of $I(t)$ affects the maximum likelihood (ML) parameter estimates; note such misspecification changes the likelihood by changing g for a given f. Since the size of the population from which $I(t)$ is estimated will generally be much larger than that of the study cohort being analyzed, one would presume that the variance in parameter estimates due to random or sampling errors in $I(t)$ would be minimal compared with the variance in the parameter estimates due to sampling errors in the study data. Note this latter variance is summarized in the parameter confidence intervals. Indeed, simulations have shown this to be the case. For example, for the data shown in Table 1 and discussed in Section 5, the variance (of the parameter estimates) due to sampling errors in $I(t)$, assuming a population of one million observed for 5 years, is only about 10–15% of the variance due to sampling errors in the study data.

There may also be some bias in the estimate of $I(t)$. Simulations we have done have shown that the effect of an upward or downward bias in $I(t)$ is predictable. Over a large range of I and f functions, using a 10 or 20% inflated (deflated) $I(t)$ always resulted in an upward (downward) bias in the estimate of sensitivity and a downward (upward) bias in the estimate of the mean of f. A 10% upward bias in $I(t)$ resulted on average in about a 15% decrease in the estimated mean of f and in about a 10% increase in estimated sensitivity; a 10% downward bias in $I(t)$ resulted in, on average, about a 25% increase in the mean of f and a 20% decrease in sensitivity.

4. Sojourn and Lead Time, Length Bias, and Overdiagnosis in the Presence of Competing Mortality

To properly evaluate various quantities related to screening, it is essential to consider competing mortality. We assume here that cumulative mortality from other causes, $M(t)$, is independent of the incidence of the disease of interest. Then the clinical and preclinical incidence probability density functions are now given by $I(t)[1 - M(t)]$ and $g(t)[1 - M(t)]$, respectively, and the sojourn time density by $f(x \mid s)[1 - M(s + x)]/[1 - M(s)]$. Note that the above densities, and others described later in this section, are improper in that they will not generally integrate to one, reflecting the fact that not all subjects will have, e.g., a time of preclinical incidence. Adding competing mortality will not change the likelihood since the incidence and prevalence rates are unchanged; however, it will affect the computation of various quantities related to the experience of a population over time, e.g., average sojourn time, average lead time, and overdiagnosis.

4.1 *Sojourn Time*

Sojourn time is usually defined as the interval from the onset of the preclinical period until the onset of clinical symptoms. In the presence of competing mortality, however, some persons with preclinical disease will never develop clinical symptoms (even with no intervention). Sojourn time could be considered censored for these individuals, or not defined, the option we choose here. In Figure 1 then, subjects 1 and 3 have sojourn times of $(T_C - T_0)$ while the sojourn time for subject 2 is not

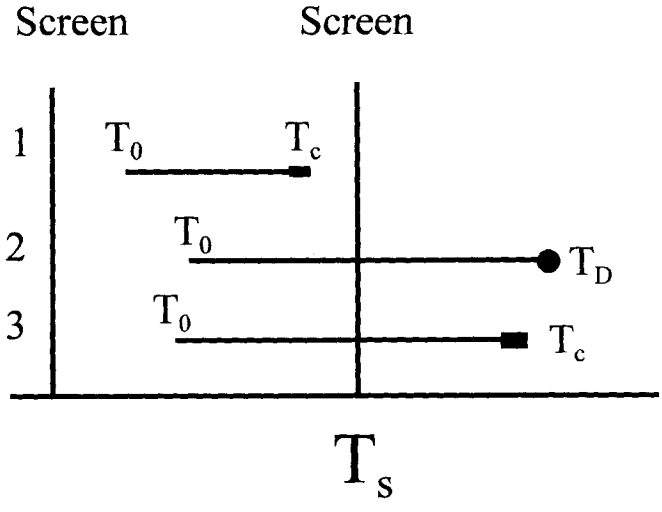

Figure 1. Schematic of preclinical disease and screening. Horizontal lines represent theoretical time courses of preclinical disease (with onset at time T_0) in three sample subjects; vertical lines are times of (100% sensitive) screening. Rectangles denote onset of clinical disease (at time T_C) while circles represent death (at time T_D).

defined. It is straightforward to show that the sojourn time density $f^*(z)$ then is given by

$$f^*(z) = \frac{\int_0^{t^F} g(s)f(z \mid s)(1 - M(s+z))ds}{\int_0^{t^F} I(t)(1 - M(t))dt}, \quad (5)$$

where t^F is an age such that $M(t^F) = 1$. The proportion of preclinical cases that ever become clinical (in the absence of screening) is then given by

$$P_{Clin} = \frac{\int_0^{t^F} I(t)(1 - M(t))dt}{\int_0^{t^F} g(t)(1 - M(t))dt}. \quad (6)$$

4.2 *Lead Time*

Lead time is the difference between the time a subject is diagnosed through screening and the time the subject would have been diagnosed without screening. The concept of program lead time examines the distribution of lead times for a population under a given screening program (Day and Walter, 1984). Incident cases are normally assigned a lead time of zero since their diagnosis is not advanced at all. Since lead time describes the advance in diagnosis among subjects who would have clinically presented (in the absence of screening), screen-identified cases who would never have presented should logically be excluded when determining the program lead time distribution (the lead times of these persons are not clearly defined anyway). In figure 1, subject 1 has a lead time of zero while subject 3 has a lead time of $T_C - T_S$; subject 2 would be excluded when considering program lead time distribution. We then define a screen-identified case to have a lead time of $y > 0$ if the subject would have been clinically diagnosed at time y after the (true) positive screen. The denominator for the program lead time distribution is all subjects who would have been clinically diagnosed at some time. Since those identified (as true positives) at different screens are mutually exclusive, the (improper) density $w(y)$ of positive lead times is equal to $\Sigma_j w_j(y)$, where $w_j(y)$ is the probability that a subject will be identified at the jth screen and have a lead time of $y > 0$. It is straightforward to show that the $w_j(y)$ are then given by

$$w_j(y) = \sum_{m=0}^{j-2} \left[\int_{t_m}^{t_{m+1}} g(s)f(y + t_j - s \mid s)M^c(t_j + y) \times \left\{ \prod_{i=1}^{j-m-1} (1 - \beta(t_i - s)) \right\} \beta(t_j - s)ds \right] + \int_{t_{j-1}}^{t_j} g(s)f(y + t_j - s \mid s)M^c(t_j + y)\beta(t_j - s),$$

where $M^c = 1 - M$ (recall the t_j are the ages at the jth screening and $t_0 = 0$). The proportion of the population having a lead time of zero, i.e., the incident cases, is given by $p(y = 0) = \int_0^{T_F} J(t)M^c(t)dt$.

4.3 *Length Bias*

Length bias refers to the tendency for screening to preferentially pick up slower growing cancers or those cancers that have a relatively long preclinical period. Denoting by x the (theoretical) length of the preclinical period in the absence of screening, we let $v(x)$ be the (improper) density of x among screen-identified cases. A measure of length bias is the mean of v divided by $\mathrm{E}(f^*)$, the average sojourn time. Similar to the distribution w of positive lead times discussed above, $v(x) = \Sigma_j v_j(x)$, where $v_j(x)$ is the probability that a subject will be identified at the jth screen and have a preclinical length of x. A subject screen diagnosed at t_j will have a (theoretical) preclinical duration of x if preclinical onset was at time $s < t_j$ and clinical diagnosis or death without prior clinical diagnosis would have occurred at time $s + x > t_j$,

$$v_j(x) = \sum_{m=0}^{j-2} \int_{t_m}^{t_{m+1}} g(s) \left\{ f(x \mid s)M^c(x+s) + M'(x+s)F^c(x \mid s) \right\} \times \left\{ \prod_{i=1}^{j-m-1} (1 - \beta(t_i - s)) \right\} \times \beta(t_j - s)I_{[x+s-t_j]}ds, + \int_{t_{j-1}}^{t_j} g(s)\{f(x \mid s)M^c(x+s) + M'(x+s)F^c(x \mid s)\} \times \beta(t_j - s)I_{[x+s-t_j]}ds,$$

where $F^c = 1 - F$ and $I_{[x]} = 1$ for $x > 0$ and is zero otherwise. Because of overdiagnosis (see below), the length bias will not approach one as the screening frequency increases.

4.4 *Overdiagnosis*

Overdiagnosis is the tendency for screening to identify cases that would never clinically present. Overdiagnosis can be quantified in two ways. We define the cohort overdiagnosis ratio (COR) as the expected number of cases that would be identified in a cohort undergoing a given screening program over the expected number of cases that would have clinically presented under no screening. In Figure 1, all three subjects would be identified as cases under the displayed screening program but only subjects 1 and 3 would have clinically presented without screening. It follows then that

$$\mathrm{COR} = \frac{\sum_{i=1}^{k} P_k M^c(t_k) + \int_0^{t^F} J(t)M^c(t)dt}{\int_0^{t^F} I(t)M^c(t)dt}. \quad (7)$$

COR can be thought of as a measure of the overall public health impact of the screening program in terms of the number of diagnosed cases. As screening becomes more frequent and more sensitive, the proportion of preclinical cases that are screen detected approaches one; thus, COR approaches the upper limit of $1/P_{\mathrm{clin}}$.

Overdiagnosis can also be conceptualized on the individual level by considering the probability that a screen-identified case would have clinically presented in the absence of screening. Note that, in Figure 1, of the two persons (2 and

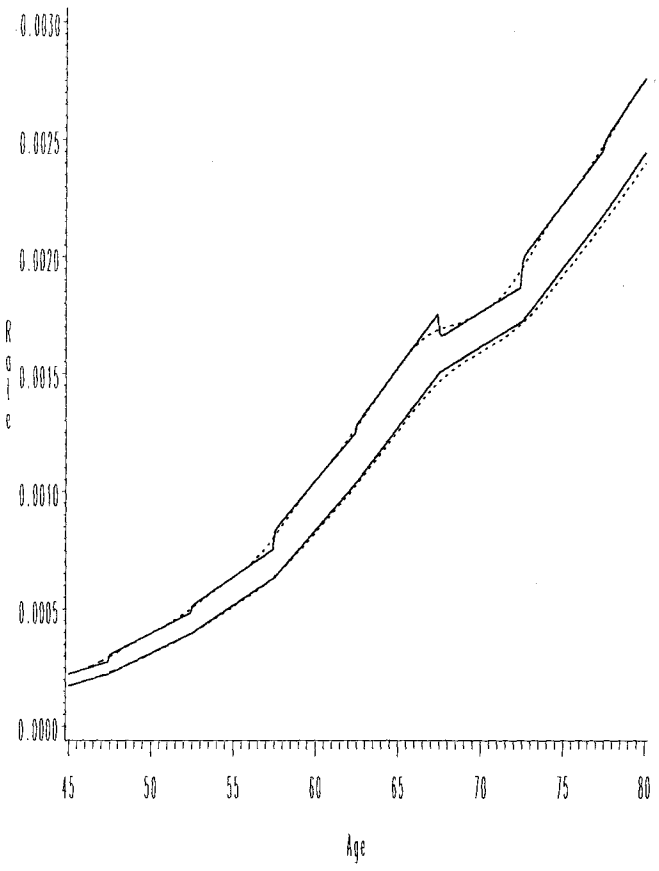

Figure 2. The lower dashed line is the reported colorectal cancer incidence curve $I(t)$ (under no screening) for the department of Calvados, France. The upper solid and dashed lines are, respectively, the unsmoothed and smoothed preclinical incidence curve g derived from the above $I(t)$ and the function f as fitted to the data in Table 1. Here f was exponential with parameter λ varying with age (see text). The lower solid line is the calculated $I(t)$ derived by convoluting the smoothed version of g with f.

3) identified through screening, only person 3 would have clinically presented. This probability can be thought of as a positive predictive value (PPV) in the following sense. Recall that PPV refers to the proportion of individuals with a positive test who are true positives. If we consider a true positive as an individual with preclinical disease who would have clinically presented in the absence of screening and consider a positive test as a positive on both the conventional screening test and on the gold standard diagnostic test (e.g., biopsy), then it is clear that the above probability is exactly the PPV. Based on the discussion of lead time above, the PPV for a case screen identified at t_j is just the probability that the case has positive lead time; hence, $\text{PPV}(t_j) = [\int_0^{t^F} w_j(y)dy]/[P_j M^c(t_j)]$.

5. Example: Community Trial of Colorectal Cancer Screening

Here we utilize data from the first round of a community screening trial of the fecal occult blood test for early detection of colorectal cancer that took place in the French department of Calvados between 1991 and 1994 (Launoy et al., 1997; Prevost et al., 1998). Briefly, a single screen was applied to the community and the cases identified at that screen, as well as incident cases in three 1-year periods postscreening, were enumerated for three different age groups (see Table 1). We derived the incidence function $I(t)$ from colorectal cancer incidence data for the department of Calvados during the period 1978–1990 when no screening was in effect (Launoy et al., 1997).

The simplest convolution model employs exponential f with constant λ and constant $\beta(t-s)$; fitting this model to the above data gave ML parameter estimates of $1/\hat{\lambda} = 4.55$ years (95% CI (3.2, 8.0)) and $\hat{\beta} = 0.50$ (95% CI (0.29, 0.70)). Using a gamma or Weibull distribution for f or assuming sensitivity was of the form $\beta(t-s) = \beta_1[(t-s)/(t-s+\beta_2)]$ did not significantly improve the fit over the two-parameter model. We also examined a model in which $f(t-s \mid s)$ was exponential with nonconstant $\lambda(s)$. Specifically, we let $1/\lambda(s) = 1/\lambda_0$ for $s \leq 40$ and $1/\lambda(s) = 1/\lambda_0 + \theta(s-40)$ for $s \geq 40$. This model had a borderline significantly better fit than did the constant λ model ($p = 0.045$, likelihood ratio test). The ML parameter estimates for this model were $\hat{\beta} = 0.44$, $1/\hat{\lambda}_0 = 1.8$ years, and $\hat{\theta} = 0.16$/year; thus, subjects with preclinical onset at ages $s = 50$ and $s = 65$ would have values of $1/\lambda$ of 3.4 and 5.8 years, respectively. Figure 2 shows the curve $I(t)$ for the Calvados department (before screening) along with the preclinical incidence function g derived with the above f (both the smoothed and unsmoothed g are displayed). The chi-square goodness-of-fit test for this model demonstrated an adequate fit to the data ($p = 0.10$); the goodness-of-fit p-value for the constant λ model was 0.03 (see Table 1).

Often, one is interested in estimates of overdiagnosis, program lead time, etc., under a proposed screening schedule that differs from the one employed in the trial being analyzed. Let $\hat{f}$ and $\hat{\beta}$ denote f and β evaluated at the ML parameter estimates of the model as fit to screening trial data. For a given proposed screening schedule, defined by the ages $t_1 \cdots t_k$ of screening, the ML estimates of the quantities described in Section 4 are then given by the corresponding formulas evaluated at $\hat{f}$ and $\hat{\beta}$, with g being derived as a function of $\hat{f}$ and $I(t)$. Profile confidence limits can be computed using the grid method (Day and Walter, 1984).

Table 2 displays ML estimates and confidence limits for various quantities under three proposed screening schedules. To facilitate comparison with traditional estimates, we utilized the model with constant λ here. The mean sojourn time (among those ever presenting) was 3.45 years, somewhat less than the mean of f of 4.5 years, and P_{clin} was equal to 0.79. We considered hypothetical screening programs that screened every 5 years, every 2 years, or every year beginning at age 50 and ending at age 80. As seen in the table, average program lead increased from 0.9 to 1.54 to 2.0 years as the frequency of screening increased from every 5 years to biannually to annually. The average lead time among screen-identified cases was relatively unaffected by screening frequency and was slightly higher than the average sojourn time. The PPV was also essentially independent of screening frequency but was dependent on age; the PPV (with annual screening) for cases screen-identified at age 50 was 0.94 compared with a PPV for cases identified at age 75 of 0.79.

Table 2
Sojourn time, lead time, length bias, and overdiagnosis

Quantity	Estimate (95% CI)	Screen frequency		
		Yearly	Every second year	Every fifth year
Mean sojourn time	3.45 (2.6, 4.9)			
P_{clin}	0.78 (0.70, 0.83)			
Mean program lead time		2.01 (1.5, 2.6)	1.54 (1.1, 2.0)	0.90 (0.70, 1.2)
Mean lead time[a]		3.66 (2.9, 5.0)	3.63 (2.9, 4.9)	3.56 (2.8, 4.9)
Length bias ratio		1.44 (1.40, 1.49)	1.61 (1.57, 1.65)	1.83 (1.78, 1.88)
Cohort overdiagnosis ratio		1.10 (1.07, 1.17)	1.08 (1.06, 1.13)	1.05 (1.03, 1.08)
PPV(50)		0.95 (0.93, 0.96)	0.95 (0.93, 0.96)	0.95 (0.93, 0.96)
PPV(75)		0.77 (0.65, 0.84)	0.77 (0.65, 0.84)	0.77 (0.65, 0.84)

[a] Among screen-detected cases.

The cohort overdiagnosis ratio decreased from 1.1 at annual screening to 1.05 for screening every 5 years, while length bias ratio increased from 1.44 with annual screening to 1.83 for screening every 5 years.

The data in Table 1 have been used previously by Launoy et al. (1997) to estimate mean sojourn time and sensitivity time using constant g and exponential f. They estimated $1/\lambda$ to be 4.9, which they reported as the mean sojourn time, and sensitivity to be 0.47.

6. Discussion

Use of the method proposed here presumes that there is a reasonable estimate of $I(t)$, the underlying incidence rate in the study population without screening. In the case of community trials, where there is generally wide participation in screening, the incidence rates for a prior period of time or the current rates in an adjacent geographic entity should give good approximations to $I(t)$. In randomized trials, study subjects are in general a more selective group, and their disease rates may not be reflective of overall local or national rates. One approach that can be used for these trials is to apply the initial choice of $I(t)$ to the (unscreened) control group to get the expected number of incident cases; a goodness-of-fit test can then be employed to test whether the chosen $I(t)$ is compatible with the observed data. Note that this approach assumes there is relatively minimal contamination (i.e., screening) in the control group. Our results on the robustness of the parameter estimates are also relevant here. We have shown here that a moderate level (e.g., 10%) of bias in $I(t)$ will not affect the parameter estimates too greatly and the direction of the effect will be predictable if the direction of bias is known. Further, we have shown that variation in parameter estimates due to random fluctuations in $I(t)$ will generally be negligible compared with the variation in parameter estimates due to random sampling error in the study data.

Acknowledgement

I thank Stuart Baker and Grant Izmirlian for helpful comments.

Résumé

Le modèle qui spécifie que l'incidence d'un cancer, I, est la convolution d'une incidence pré clinique, g, et de la densité de temps dans la phase pré clinique, f, a été fréquemment utilisé pour modéliser des données d'essai de protection de cancer et pour estimer des quantités telles le temps de séjour, le temps restant et la sensibilité. Quand ce modèle est ajusté à ces données, les paramètres de f aussi bien que ceux qui gouvernent la sensibilité à la protection doivent être estimés. Auparavant, g était soit supposé égal à l'incidence clinique ou supposé constant ou une fonction exponentielle qui devait aussi être estimé. Ici nous supposons que l'incidence sous-jacente I dans la population étudiée (en l'absence de protection) est connue. Avec I connue, g devient une fonction de f dont on peut avoir la solution en effectuant une déconvolution (numérique), éliminant ainsi le besoin de l'estimer ou de l'assortir de suppositions. Puisque les procédures de déconvolution numérique sont néanmoins hautement instables, nous incorporons une procédure de lissage qui produit une fonction g réaliste tandis qu'elle reproduit de très près la fonction d'incidence initiale I dans sa convolution avec f. Nous ajoutons aussi le concept de mortalité compétitive au modèle de convolution. Ainsi, de conserve avec la fonction d'incidence pré clinique réaliste décrite ci-dessus, on aboutit à des estimations plus précises de temps de séjour et de temps restant, et on obtient une estimation de quantités reliées à un diagnostic excessif que nous définissons ici.

References

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.

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.

Duffy, S. W., Chen, H. H., Tabar, L., and Day, N. E. (1995). Estimation of mean sojourn time in breast cancer screening using a Markov chain model of both entry to and exit from the preclinical detectable phase. *Statistics in Medicine* **14,** 1531–1543.

Launoy, G., Smith, T. C., Duffy, S. W., and Bouvier, V. (1997). Colorectal cancer mass-screening: Estimation of faecal occult blood test sensitivity, taking into account cancer mean sojourn time. *International Journal of Cancer* **73,** 220–224.

Liao, J. and Brookmeyer, R. (1995). An empirical Bayes approach to smoothing backcalculation of HIV infection rates. *Biometrics* **51,** 579–588.

Linz, P. (1985). *Analytical and Numerical Methods for Volterra Equations.* Philadelphia: SIAM.

O'Sullivan, F. (1986). A statistical perspective on ill-posed inverse problems. *Statistical Science* **1,** 502–527.

Prevost, T. C., Launoy, G., Duffy, S. W., and Chen, H. H. (1998). Estimating sensitivity and sojourn time in screening for colorectal cancer. *American Journal of Epidemiology* **148,** 609–619.

Ries, L. A. G., Kosary, C. L., Hankey, B. F., Miller, B. A., Clegg, L., and Edwards, B. K. (1999). *SEER Cancer Statistics Review, 1973–1996.* Bethesda, Maryland: National Cancer Institute.

Shen, Y. and Zelen, M. (1999). Parametric estimation procedures for screening programmes: Stable and nonstable disease models for multimodality case finding. *Biometrika* **86,** 503–515.

Simonoff, J. B. (1996). *Smoothing Methods in Statistics.* New York: Springer.

Received February 2000. *Revised October* 2000.
Accepted October 2000.

Appendix

We use as the measure of error the average value of $\mid I_h(t) - I(t) \mid /I(t)$ over the age interval $(t_a, t_b) = (20, 80)$, where $I_h(t) = \int_0^t g_h^*(s) f(t-s \mid s) ds$. For smoothness, we considered the ratio of the curvature of g_h^* with that of a pure cubic polynomial (cx^3) whose mean on (t_a, t_b) is the same as that of g_0^*; curvature was defined here as the average of the absolute value of the second derivative. The smoothing function utilized was a kernel smoothing routine with an Epanechnikov kernel, i.e., $(1-u^2)$ (Simonoff, 1996).

Over a wide range of gamma and Weibull f functions and incidence functions $I(t)$ derived from Surveillance, Epidemiology, and End Results (SEER) data (Ries et al., 1999), we found that using a bandwidth of from 1.5 to 2.0 in general gave adequate error rates as well as adequate smoothness as measured by the curvature ratio (CR). For example, for 280 combinations of (I, f) where the auxiliary gamma or Weibull parameter b (where $f(t) \propto t^{b-1}$) was 1.5 or less (note this includes exponential f where $b = 1$), the error was always less than 0.05 (average about 0.01) while the CR was always less than 10 (average about 1.0). The resulting g function was nonnegative all but 0.7% of the time. In contrast, with no smoothing ($h = 0$), the CR exceeded 100 about 13% of the time and was not nonnegative 9.3% of the time. When the parameter b was 2.0 or greater and especially when the mean of f was large ($\geq$7 years), finding a reasonable h value became problematic. At h values large enough to give reasonable smoothness, the error became larger than acceptable (>0.05). While this may be in part a numerical phenomenon, it may also indicate that such f functions are implausible since no reasonable preclinical incidence function g is compatible with such an f and the given incidence function $I(t)$.

Note that the numerical algorithm for generating g^* depends on having an essentially continuous expression for $I(t)$. Here we take (in general) the 5-year incidence rates and linearly interpolate to get a continuous function $I(t)$.