American Journal of Epidemiology

Vol. 148, No. 6
Printed in U.S.A.

Estimating Sensitivity and Sojourn Time in Screening for Colorectal Cancer

A Comparison of Statistical Approaches

T. C. Prevost,[1] G. Launoy,[2] S. W. Duffy,[1] and H. H. Chen[3]

The effectiveness of cancer screening depends crucially on two elements: the sojourn time (that is, the duration of the preclinical screen-detectable period) and the sensitivity of the screening test. Previous literature on methods of estimating mean sojourn time and sensitivity has largely concentrated on breast cancer screening. Screening for colorectal cancer has been shown to be effective in randomized trials, but there is little literature on the estimation of sojourn time and sensitivity. It would be interesting to demonstrate whether methods commonly used in breast cancer screening could be used in colorectal cancer screening. In this paper, the authors consider various analytic strategies for fitting exponential models to data from a screening program for colorectal cancer conducted in Calvados, France, between 1991 and 1994. The models yielded estimates of mean sojourn time of approximately 2 years for 45- to 54-year-olds, 3 years for 55- to 64-year-olds, and 6 years for 65- to 74-year-olds. Estimates of sensitivity were approximately 75%, 50%, and 40% for persons aged 45–54, 55–64, and 65–74 years, respectively. There is room for improvement in all models in terms of goodness of fit, particularly for the first year after screening, but results from randomized trials indicate that the sensitivity estimates are roughly correct. *Am J Epidemiol* 1998;148:609–19.

colorectal neoplasms; mass screening; models, statistical; sensitivity and specificity

BACKGROUND

Screening for occult disease can be carried out for three major purposes: to eliminate those already infected as part of the strategy of an immunization program; to identify and quarantine carriers of an infective agent; and to advance the stage of disease at diagnosis to facilitate curative treatment. Screening for cancer falls into the last category. In this case, two crucial elements are the sojourn time and the sensitivity of the screening test. The former is defined as the duration of the preclinical screen-detectable period, that period during which a person is asymptomatic but the disease is detectable by a screening tool. The latter is the probability that any given case who is subjected to the screening method during this period will have his or her disease detected by it.

For any given disease, one would expect the sojourn time to vary between cases, in that tumors grow at varying rates, depending on numerous pathologic and host factors. Therefore, the parameter estimated in practice in screening programs is the average sojourn time over all disease cases, usually referred to as the mean sojourn time. A long mean sojourn time indicates a good potential for screening. The shorter the sojourn time, the more frequently screening has to take place in order to be effective. If the mean sojourn time is very short, screening may not be worthwhile at all.

Note that this parameterization of the problem is a simplification of the biologic process. One would expect screen-detectability, as measured by sensitivity, to increase continuously with time, as the tumor grows, with varying rates of increase for different individuals. Under the mean sojourn time/sensitivity model as traditionally used, we approximate this by a screen-detectability which is zero up to the beginning of the preclinical screen-detectable period and is equal to constant sensitivity throughout the preclinical screen-detectable period. The length of the preclinical screen-detectable period is variable between individuals in this model. This is illustrated in figure 1. The sigmoid curve shows the true detectability and the rectangle the approximation which is traditionally assumed. This paper uses the rectangular assumption throughout for the sake of simplicity, and because it has previously been found to give a reasonable fit (1). However, the reader should bear in mind that it is a simplification.

Received for publication September 4, 1997, and accepted for publication February 19, 1998.
[1] Institute of Public Health, University Forvie Site, Cambridge CB2 2SR, United Kingdom.
[2] Registre des cancers digestifs du Calvados, INSERM, Avenue Côte de Nacre, 14033 Caen cedex, France.
[3] Graduate Institute of Epidemiology, College of Public Health, National Taiwan University, Taipei, Taiwan.

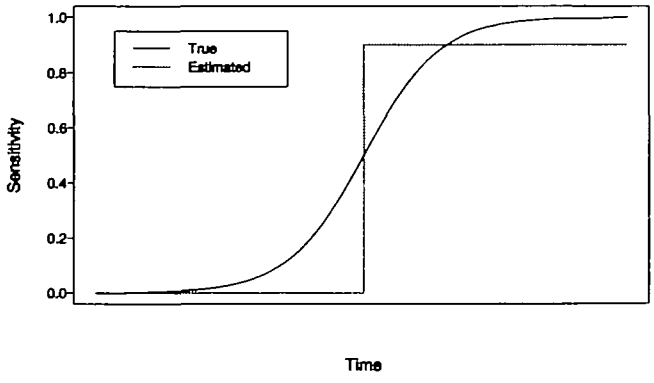

FIGURE 1. Postulated true detectability (—) of occult disease as measured by the sensitivity of a screening test and modeled detectability (· · ·), using the usual mean sojourn time/sensitivity estimation model.

A considerable body of literature on estimation of mean sojourn time and sensitivity has been built up over the years (for reviews, see Stevenson (1) and van Oortmarssen et al. (2)). The seminal work in the field was carried out by Zelen and Feinleib (3), Prorok (4), and Day and Walter (5). Mathematical modeling has been used in the context of cervical cancer (6–8), lung cancer (9), and colorectal cancer (10). Many of the applications, however, have been in the field of screening for breast cancer, partly because of the data sets available from the large number of randomized trials of breast cancer screening (11). The major requisite of estimation is data on screening for the disease, including data on interval cancers, those which are diagnosed clinically after a negative screen. Clearly, if screening is sensitive and if mean sojourn time is reasonably long, there should be relatively few such cancers. Observation of the rate at which the incidence of interval cancers approaches the incidence observed in the absence of screening is essential to determination of the mean sojourn time.

Much of the work carried out in the past has involved exponential models of time to clinical disease, since Walter and Day have shown it to give a good fit to breast cancer screening data (12). In addition, it is a mathematically easy distribution with which to work, fitting in well with available Poisson regression computer programs (13). This is relevant, because in this field the available data often dictate that simplifying assumptions be made. These include distributional forms like the exponential, the assumption of a constant sensitivity and a single parameter for mean sojourn time (in turn necessitating several analyses in different strata), and the use of a fixed uniform underlying incidence estimated from randomized or historical control data (3, 5, 13).

A technique of estimation which permits analysis of data of complex structure and multiparameter estimation from such data is Gibbs sampling (14). The technique has been used in modeling cervical cancer screening data (15), and it seems appropriate to extend its use to models for interval cancers in screening for tumors at other sites, particularly to estimation of mean sojourn time and sensitivity.

Colorectal cancer screening using fecal Hemoccult testing (Smithkline Corporation, Philadelphia, Pennsylvania), in conjunction with colonoscopy upon positive Hemoccult results, has been shown to be effective in randomized trials (10–12). There is little information in the literature on the development of the disease and the estimation of the sojourn time and sensitivity as pointers to the potential for early detection. It would be of some interest to demonstrate whether methods commonly used in breast cancer screening could be used in colorectal cancer screening. In this paper, we consider various analytic strategies, including novel estimation strategies based on Gibbs sampling using the program BUGS (16). We fit exponential models to data obtained from a screening program for colorectal cancer conducted in Calvados, France (17).

DATA

A single round of mass screening for colorectal cancer using fecal Hemoccult testing was conducted in Calvados, France, between April 1991 and May 1994. A total of 71,307 people between the ages of 45 and 74 years were invited to undergo screening. Of the 2,020 persons with a positive Hemoccult test result, colonoscopy was performed on 1,603, and 131 cancers were detected. For a 3-year period following the screening, any cancers occurring in the negatively screened subjects were recorded along with the number of person-years contributed to follow-up. The expected incidence in the absence of screening was estimated by the incidence rates provided by the digestive cancer registry for people aged 45–74 years in the same area during 1978–1990—i.e., the period just before screening (table 1). Further details on the study can be found in the paper by Launoy et al. (17). For assessment of the potential effectiveness of Hemoccult screening for colorectal cancer in this population, the sensitivity of the Hemoccult test and the mean sojourn time of colorectal cancer are useful indicators.

Below we describe methods for obtaining estimates of screening test sensitivity and mean sojourn time based on modeling of the incidence of interval cancers. Throughout, a two-stage disease process is assumed in which an asymptomatic but screen-detectable preclinical phase precedes clinical disease.

TABLE 1. Numbers of colorectal cancers in the control population (1978–1990), screen-detected cancers, and interval cancers in a screening program conducted in Calvados, France (1991–1994), by age at detection

Age (years) at detection	Controls		At screening		First year after screening		Second year after screening		Third year after screening	
	No. of cancers	Person-years of observation	No. of cancers	Person-years of observation	No. of cancers	Person-years of observation	No. of cancers	Person-years of observation	No. of cancers	Person-years of observation
45–54	249	799,303	12	25,041	4	24,545	3	22,568	2	14,314
55–64	612	747,773	35	24,933	16	24,211	7	22,456	8	14,494
65–74	842	531,858	84	19,836	24	19,068	19	17,643	12	11,045
Unknown			0	1,497						
Total	1,703	2,078,934	131	71,307	44	67,824	29	62,667	22	39,853

MODELS

The observed incidence of cancers in the interval $(t - 1, t]$ years after a negative screen, r_t, is assumed to follow a Poisson process, with an expected value which depends on t, the time since screening, λ, the rate at which preclinical disease progresses to clinical disease, and S, the sensitivity of the screening test. This can be expressed as

$$r_t \sim \text{Poisson}(I(t, \lambda, S)),$$

with $I(t, \lambda, S)$ being the expected incidence of cancers in the interval $(t - 1, t]$ years after a negative screen. Hence, the likelihood for λ and S, given the observed incidence of interval cancers within each yearly interval from the time of screening (t_0) to k years later, is

$$\prod_{i=0}^{k} \frac{I(t_i, \lambda, S)^{r_{t_i}} e^{-I(t_i,\lambda,S)}}{r_{t_i}}.$$

Before we can proceed with the estimation of λ and S, some functional form for the expected incidence $I(t, \lambda, S)$ needs to be specified. In the following sections, we develop and compare three different expressions for $I(t, \lambda, S)$.

Model 1—incidence of interval cancers modeled as a function of mean sojourn time

Here we assume sensitivity to be 100 percent in the first instance. When clinical disease occurs t years after a screening in a person correctly screened negative, two inferences about the natural history of the disease will be made: 1) preclinical disease has arisen since the screening and 2) the duration of the preclinical screen-detectable period was at most t years. The probability of developing clinical disease in the interval $(t - 1, t]$ after a negative screen is equal to the probability that the sum of the time from screening to entering the preclinical phase and the sojourn time lies between $t - 1$ and t. This is the probability that the sojourn time s, let us say, does not exceed t and the time of "birth" into the preclinical phase occurs between $t - s - 1$ and $t - s$ for values of s less than $t - 1$, and between 0 and $t - s$ for values of s greater than $t - 1$. This means that if incidence of birth into the preclinical period is uniform over the course of a year, the expected incidence lies between

$$J\int_0^{t-1} f(s)ds$$

and

$$J\int_0^t f(s)ds,$$

where J denotes the incidence of preclinical disease and $f(y)$ is the probability density function of the duration of the preclinical screen-detectable period. Typically, the incidence of cancers in an unscreened control population is taken as an estimate of the incidence of preclinical disease, J.

Depending on the form of $f(s)$, the exact formula may be complex. A possible approximation is to define t' as $t - 0.5$ and estimate expected incidence as

$$I(t) = J\int_0^{t'} f(s)ds. \quad (1)$$

This approximation is implicitly used by Paci and Duffy (13).

If an exponential distribution with parameter λ is assumed for the duration of the preclinical screen-detectable period, thereby implying that the transition rate from preclinical disease to clinical disease is constant over time, equation 1 becomes

$$I(t, \lambda) = J \times (1 - e^{-\lambda(t-0.5)}), \quad (2)$$

where λ represents the rate at which preclinical disease progresses to clinical disease and $1/\lambda$ provides an estimate of the mean sojourn time.

Paci and Duffy (13) specify the incidence of interval cancers as a Poisson process with the mean as given in equation 2, and they use generalized linear modeling techniques to obtain an estimate of λ (and hence the mean sojourn time), based on the observed incidence of interval cancers. The sensitivity of the screening test can then be calculated as a function of the estimated transition rate, $\hat{\lambda}$. A formula for estimating test sensitivity in this way is given by Paci and Duffy. This formed the basis for the estimate derived from Markov chain modeling by Duffy et al. (18). We use the latter estimate here. The formula is

$$\hat{s}(t) = 1 - [K(t)(\lambda - J)/n - a]/b, \quad (3)$$

where

$$a = (\lambda - J)e^{-JT}(1 - e^{-Jt}) - Je^{-\lambda t - JT}(e^{(\lambda - J)t} - 1)$$

and

$$b = J(e^{-JT} - e^{-\lambda T})(1 - e^{-\lambda t}).$$

$K(t)$ is the cumulative number of cases t years after a negative screen and T is age at the time of screening. In this paper, sensitivity, S, has been estimated as a weighted average of $\hat{s}(t)$ for $t = 1$, 2, and 3, using $K(t)$ as the weights.

Model 2—Incidence of interval cancers and screen-detected cancers modeled as a function of sensitivity and mean sojourn time

If the sensitivity of the screening test is substantially less than 100 percent, then the expected incidence of interval cancers will not be adequately described by equation 2. The cases of preclinical disease which go undetected at the screening, in the absence of further screening, will at some point in time surface as clinical cancers, and more cancers than expected will be observed in the years following screening.

When clinical disease occurs t years after a screening in a person incorrectly screened as negative, it will be inferred that 1) preclinical disease was present at the time of screening and 2) the duration of the preclinical screen-detectable period was at least t years. Extending equation 2 to account for cancers arising in individuals who were falsely screened negative gives the expected incidence of interval cancers t years after screening as

$$I(t, \lambda, S) = J \times [1 - \exp(-\lambda(t - 0.5))] + \frac{c(1 - S)}{S}[\exp(-\lambda(t - 1)) - \exp(-\lambda t)], \quad (4)$$

where c is the number of cancers detected at the screening and S is the test sensitivity. $[c(1 - S)]/S$ is the expected number of cancers which were present but not detected at screening, and $\exp(-\lambda(t - 1)) - \exp(-\lambda t)$ is the probability that a cancer which went undetected at screening will surface clinically in the interval $(t - 1, t]$.

In addition to the incidence of interval cancers, data on the prevalence of preclinical disease detected at screening are also available. Based on the Markov chain model of Duffy et al. (18), the expected prevalence can be expressed in terms of λ and S as

$$P(\lambda, S) = \frac{n_s SJ(e^{-\lambda T} - e^{-JT})/(J - \lambda)}{e^{-JT} + J(e^{-\lambda T} - e^{-JT})/(J - \lambda)}, \quad (5)$$

where T is age at the time of screening and n_s is the number of people screened.

Incorporating this additional information into the modeling by assuming that the number of cancers detected at screening, c, is binomial $(n_s, P(\lambda, S)/n_s)$ will improve the estimation of λ and S. The likelihood for λ and S then becomes

$$\frac{n_s!}{c!(n_s-c)!}\left[\frac{P(\lambda, S)}{n_s}\right]^c \left[1-\frac{P(\lambda, S)}{n_s}\right]^{(n_s-c)} \prod_{i=0}^{k} \frac{I(t_i, \lambda, S)^{r_{t_i}} e^{-I(t_i,\lambda,S)}}{r_{t_i}}. \quad (6)$$

Model 3—incidence of interval cancers and screen-detected cancers modeled as a function of sensitivity and mean sojourn time, using exact times of transition

When they are known, exact times at which interval cancers surface clinically can be used in place of yearly incidences of interval cancers. In their modeling of exact times of interval cancer occurrence, Duffy et al. (18) found that using exact time to occurrence improved the precision of estimation. The probability of clinical disease at exactly t years for those correctly screened negative is given as

P(no disease up to $t - \delta t$ years)

$\times P$(no disease to clinical in interval $(t - \delta t, t)$)

$+ P$(no disease to preclinical in interval $(0, t - \delta t)$)

$\times P$(preclinical to clinical in interval $(t - \delta t, t)$).

For those falsely screened negative, the probability of clinical disease at exactly t years is

$$\frac{c(1-S)}{S} \times P(\text{preclinical up to } t - \delta t \text{ years}) \times P(\text{preclinical to clinical in interval } (t - \delta t, t)).$$

δt is taken to be the limit of accuracy, here 1 month ($\approx$0.08 years).

Therefore, the expected monthly incidence of interval cancers, $I(t, \lambda, S)$, can be expressed as

$$py_i\left\{e^{-J(t_i-0.08)}\left(1+\frac{\lambda e^{-0.08J}-Je^{-0.08\lambda}}{J-\lambda}\right) + \frac{J(e^{-\lambda(t_i-0.08)}-e^{-J(t_i-0.08)})(1-e^{-0.08\lambda})}{J-\lambda}\right\} + \frac{c(1-S)}{S}(e^{-\lambda(t_i-0.08)}-e^{-\lambda t_i}),$$

where py_i is the number of person-years contributed in the yth year after screening, with y such that $y - 1 < t_i \leq y$.

The likelihood for λ and S is then given as in equation 6, except that the interval $(t_{i-1}, t_i]$ is 1 month, and r_{t_i} is the observed monthly incidence of cancers after a negative screen.

PARAMETER ESTIMATION

Two approaches can be applied to the estimation of λ and S. The first involves expressing test sensitivity in terms of λ (as in the formula given in equation 3) and hence reduces the problem to estimation of one parameter only. The second would be to attempt to estimate λ and S as independent parameters in a single likelihood, without specifying a direct relation between them. Ideally, the second approach would be preferred, since it does not rely on any prespecified, and potentially incorrect, relation between mean sojourn time and sensitivity. However, in practice, two-parameter estimation may be difficult because of the small amounts of data typically available and the inherently high correlation between estimates of sensitivity and mean sojourn time.

The Paci and Duffy model (model 1) is a one-parameter model. However, models 2 and 3 could be treated as having either one or two unknown parameters. In this paper, λ and S are estimated independently in these models and the correlation between the reported estimates is given.

In applying the three models described above to the Calvados colorectal cancer screening data, several estimation techniques were used: maximum likelihood, a least squares approach, and Gibbs sampling.

Maximum likelihood

Maximization of the likelihood was performed over values of λ in the one-parameter model and over values of λ and S in the two-parameter models. Variance estimates were obtained using the inverse Fisher information matrix. The calculations were performed using Mathematica (19).

Least squares

The SAS NLIN procedure (20) was performed on the nonlinear regression of the observed outcomes on the expected outcomes as given by $I(t, \lambda, S)$ and $P(\lambda, S)$, with λ and S specified as regression coefficients to be estimated.

Gibbs sampling

A Bayesian approach to estimation in which both data and parameters are assumed to be random variables can be taken. The joint probability distribution for the unknown parameters and the data is obtained by multiplying the likelihood function by prior probability distributions for the unknown parameters. From this joint probability distribution, the posterior distri-

bution—that is, the distribution for the unknown parameters conditional on the observed data—can be derived. Simulated values from the marginal posterior distribution for any parameter of interest can be obtained using Gibbs sampling implemented by the BUGS software (16). From the sample of simulated values obtained, empirical estimates of summary statistics (e.g., means, medians, standard deviations) for the parameters of interest are calculated.

A gamma(0.001,0.001) distribution truncated outside the region 0.05–5 was taken as the prior for λ and a uniform(−4,4) distribution as the prior for log ([1−S]/S). Convergence was assessed using the method of Gelman and Rubin (21). The BUGS model specification code for model 2 is given in the Appendix.

RESULTS

Overall, the three estimation techniques gave very similar estimates of λ and S for models 2 and 3 (tables 2–4). Models 2 and 3 yielded mean sojourn times which increasd with age and sensitivity estimates which declined with age. For the age group 45–54 years, mean sojourn time estimates were approximately 2 years and sensitivity estimates were approximately 75 percent; for the age group 55–64 years, the corresponding figures were 3 years and 50 percent; and for the age group 65–74 years, the estimates were 6 years and 40 percent. Model 1 estimated uniformly short mean sojourn times and high sensitivities. The 95 percent intervals obtained from maximum likelihood and Gibbs sampling agreed well, but the least squares intervals sometimes disagreed.

One characteristic common to models 2 and 3 is the high positive correlation between estimates of λ and S (in model 1, S is by definition a function of λ). Plots of the 95 percent confidence regions for the maximum likelihood estimates from model 2 (figure 2) and model 3 (figure 3) clearly show the high correlation and the large degree of uncertainty associated with the estimates, particularly in the youngest age group, where the number of observed interval cancers was small.

Model 1 did not fit the data well (figures 4–6), which was to be expected. The model assumes that any interval cancers are newly arising cancers not present at the time of screening. Under this assumption, we would observe an increasing incidence of interval cancers with increasing time since screening. For the Calvados data presented here, this is clearly not the case. In each age group, the highest incidence was found in the year immediately following screening. Thus, model 1 tends to overestimate later interval cancers because of the high observed rates of earlier cancers. This high incidence in the first year may be due to low sensitivity to certain tumors and/or a short mean sojourn time, in which case the incidence of interval cancers in the second and third years would be expected to be similar to the control population incidence. However, the incidence of interval cancers in the second and third years in all of the age groups was well below the control population incidence.

DISCUSSION

Figures 4–6 show that models 2 and 3, as expected, gave a slightly better fit than model 1, particularly in the older age groups, where incidence was higher. There is room for improvement in all models and all estimation techniques in terms of goodness of fit. The

TABLE 2. Estimation of the preclinical-to-clinical disease transition rate, λ, and fecal Hemoccult test sensitivity, S, using various techniques applied to data for persons aged 45–54 years in a colorectal cancer screening program conducted in Calvados, France (1991–1994)

	Estimate		Correlation between $\hat{\lambda}$ and $\hat{S}$	Residual*		
	$\hat{\lambda}$	$\hat{S}$		Year 1	Year 2	Year 3
Model 1						
Maximum likelihood	0.684 (0.20–2.29)†	0.93		−0.45	0.51	0.83
Least squares	0.486 (0.16–1.50)	0.87		−0.59	0.21	0.57
Gibbs sampling	0.688 (0.34–1.25)	0.93 (0.81–0.97)		−0.44	0.51	0.83
Model 2						
Maximum likelihood	0.422 (0.13–1.40)	0.70 (0.11–0.98)	0.92	−0.19	0.49	0.83
Least squares	0.472 (0.16–1.38)	0.75 (0.06–0.99)	0.97	−0.22	0.51	0.84
Gibbs sampling	0.478 (0.09–0.98)	0.80 (0.19–0.98)	0.78	−0.25	0.51	0.86
Model 3						
Maximum likelihood	0.439 (0.14–1.41)	0.71 (0.11–0.98)	0.91	−0.24	0.43	0.76
Least squares	0.487 (0.22–1.07)	0.75 (0.12–0.99)	1.00	−0.25	0.46	0.78
Gibbs sampling	0.487 (0.08–1.00)	0.80 (0.18–0.98)	0.78	−0.28	0.46	0.80

* Residual = (number of cancers estimated from model − observed number of cancers)/observed number of cancers.
† Numbers in parentheses, 95% confidence interval.

TABLE 3. Estimation of the preclinical-to-clinical disease transition rate, λ, and fecal Hemoccult test sensitivity, *S*, using various techniques applied to data for persons aged 55–64 years in a colorectal cancer screening program conducted in Calvados, France (1991–1994)

	Estimate		Correlation between $\hat{\lambda}$ and $\hat{S}$	Residual*		
	$\hat{\lambda}$	$\hat{S}$		Year 1	Year 2	Year 3
Model 1						
Maximum likelihood	2.337 (0.47–11.70)†	0.87		−0.15	1.55	0.48
Least squares	0.910 (0.16–5.19)	0.80		−0.55	0.95	0.33
Gibbs sampling	0.930 (0.64–1.31)	0.80 (0.73–0.84)		−0.54	0.97	0.34
Model 2						
Maximum likelihood	0.273 (0.11–0.68)	0.49 (0.20–0.79)	0.95	−0.29	0.84	0.37
Least squares	0.294 (0.07–1.20)	0.51 (0.07–0.93)	0.98	−0.30	0.84	0.36
Gibbs sampling	0.289 (0.08–0.64)	0.51 (0.17–0.90)	0.93	−0.31	0.83	0.33
Model 3						
Maximum likelihood	0.286 (0.11–0.72)	0.50 (0.19–0.81)	0.96	−0.32	0.77	0.31
Least squares	0.299 (0.16–0.56)	0.51 (0.23–0.79)	1.00	−0.32	0.78	0.31
Gibbs sampling	0.299 (0.08–0.67)	0.52 (0.17–0.91)	0.93	−0.34	0.77	0.30

* Residual = (number of cancers estimated from model − observed number of cancers)/observed number of cancers.
† Numbers in parentheses, 95% confidence interval.

main difficulty for the models lies in the very high rates of clinical cancer diagnosed in the first year after screening. One possible enhancement to the models might be differential sensitivity, whereby certain types of cancer which have a short sojourn time are more likely to be missed by screening. This seems unlikely but not impossible. There does seem to be considerable variation in sojourn times by subsite of tumor—distal colon, proximal colon, or rectum (17).

Another possibility is an increase in the reporting of early symptoms due to heightened awareness of the condition brought about by the screening. Examination of interval cancers by Dukes' stage should throw some light on this possibility, since the proportion of cancers diagnosed at the earliest stage of development is higher after a negative test, particularly in the first year, than would be expected in the absence of screening. One might therefore seek to model a lowering of the threshold for clinical detection in the period immediately after screening. In terms of formal modeling, this would entail estimation of multiple parameters relating mean sojourn time to time since screening, and problems of identifiability might arise. A simple expedient might be to fit the model in equa-

TABLE 4. Estimation of the preclinical-to-clinical disease transition rate, λ, and fecal Hemoccult test sensitivity, *S*, using various techniques applied to data for persons aged 65–74 years in a colorectal cancer screening program conducted in Calvados, France (1991–1994)

	Estimate		Correlation between $\hat{\lambda}$ and $\hat{S}$	Residual*		
	$\hat{\lambda}$	$\hat{S}$		Year 1	Year 2	Year 3
Model 1						
Maximum likelihood	2.798 (0.97–8.10)†	0.89		−0.05	0.45	0.46
Least squares	2.697 (0.73–10.03)	0.89		−0.07	0.44	0.46
Gibbs sampling	1.106 (0.83–1.43)	0.70 (0.63–0.75)		−0.47	0.19	0.36
Model 2						
Maximum likelihood	0.165 (0.08–0.32)	0.44 (0.23–0.67)	0.96	−0.22	0.05	0.47
Least squares	0.162 (0.08–0.33)	0.43 (0.19–0.71)	0.98	−0.22	0.05	0.47
Gibbs sampling	0.158 (0.06–0.29)	0.42 (0.19–0.66)	0.96	−0.21	0.06	0.48
Model 3						
Maximum likelihood	0.152 (0.07–0.31)	0.41 (0.20–0.64)	0.97	−0.22	0.04	0.47
Least squares	0.149 (0.08–0.27)	0.40 (0.20–0.64)	1.00	−0.21	0.04	0.47
Gibbs sampling	0.149 (0.06–0.27)	0.40 (0.18–0.63)	0.96	−0.21	0.05	0.48

* Residual = (number of cancers estimated from model − observed number of cancers)/observed number of cancers.
† Numbers in parentheses, 95% confidence interval.

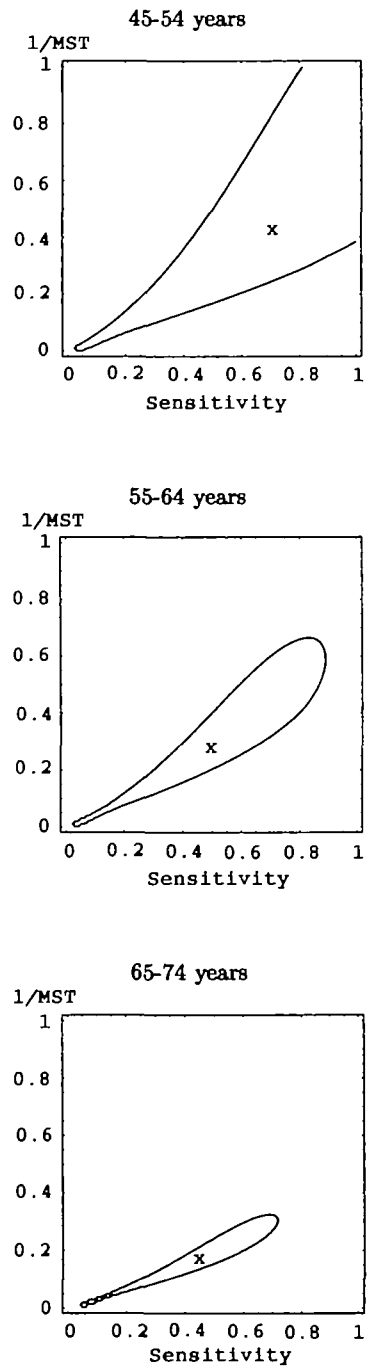

FIGURE 2. Maximum likelihood estimate (x) and 95% confidence region (loop) for 1/(mean sojourn time (MST)) and test sensitivity under model 2, using data from colorectal cancer screening in Calvados, France (1991–1994).

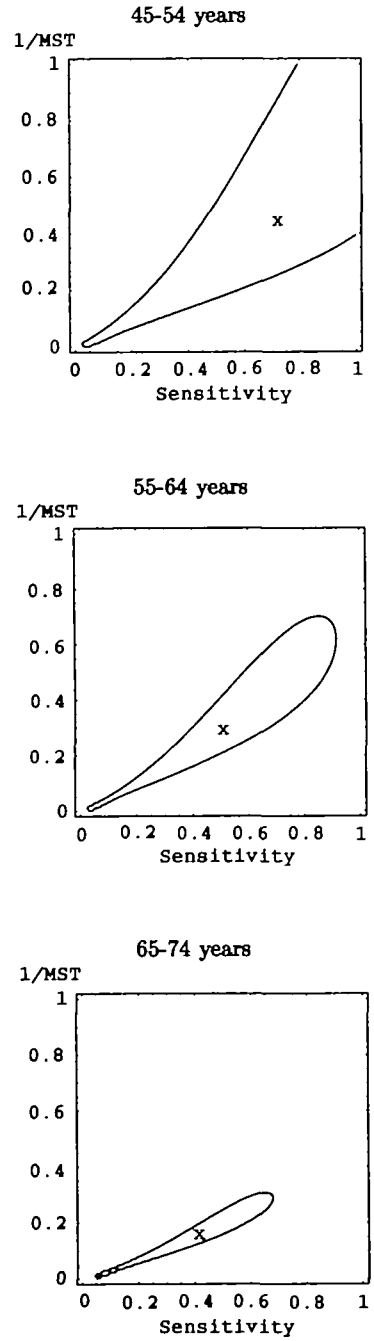

FIGURE 3. Maximum likelihood estimate (x) and 95% confidence region (loop) for 1/(mean sojourn time (MST)) and test sensitivity under model 3, using data from colorectal cancer screening in Calvados, France (1991–1994).

tion 4 without the first-year data and then estimate the difference in sojourn time required to achieve the excess observed in the first year. This model assumes a constant mean sojourn time in the first year after a negative screen, followed by a different (higher) mean sojourn time thereafter. It is certainly a feasible strategy and it may give a reasonable fit to the data, but the search for a more elegant model which would allow mean sojourn time to change continuously should continue.

The problem may also lie in the joint distribution of sojourn time and sensitivity. Clearly, a realistic model

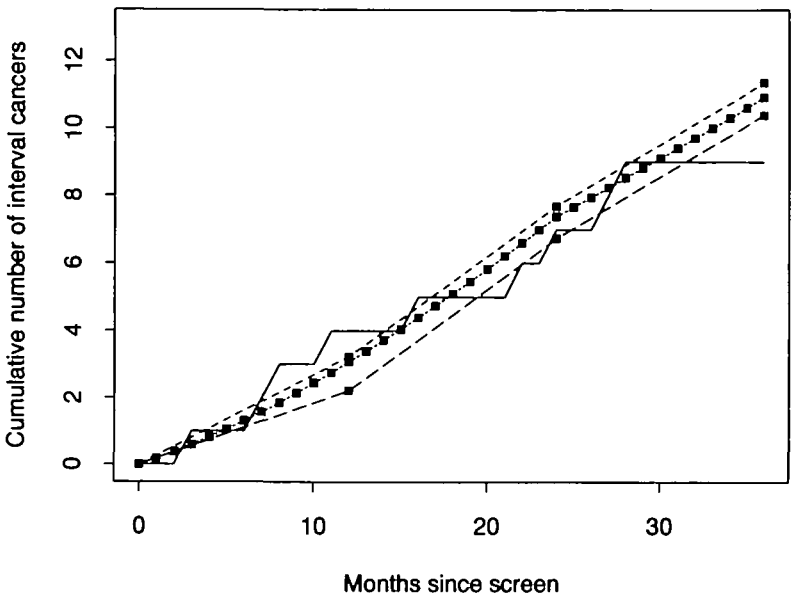

FIGURE 4. Observed (—) and maximum likelihood estimates of (model 1: – – –, model 2: ----, model 3: · · · ·) cumulative incidence of colorectal cancer following a negative screen at age 45–54 years, based on data from colorectal cancer screening in Calvados, France (1991–1994).

would be one in which sensitivity increased in a sigmoid fashion throughout the period of the tumor's growth. In our models, this sigmoid curve is approximated by a rectangle of varying lengths (sojourn time) but constant height (sensitivity). This model has proven to give a good fit to breast cancer screening data (12), but it has not been used on colorectal cancer data until now.

Similarly, the assumption of an exponential distribution of sojourn time has been shown to give a good fit for breast cancer but not as yet for colorectal cancer. It should be noted, however, that a change in the distribution of sojourn time alone is unlikely to account for the high rate of interval cancers in the first year after screening.

In the models presented here, the incidence of pre-

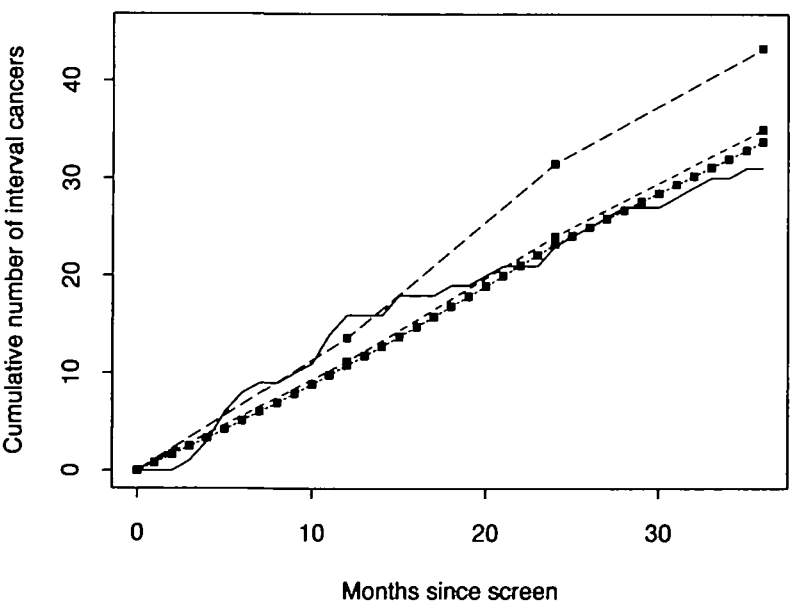

FIGURE 5. Observed (—) and maximum likelihood estimates of (model 1: – – –, model 2: ----, model 3: · · · ·) cumulative incidence of colorectal cancer following a negative screen at age 55–64 years, based on data from colorectal cancer screening in Calvados, France (1991–1994).

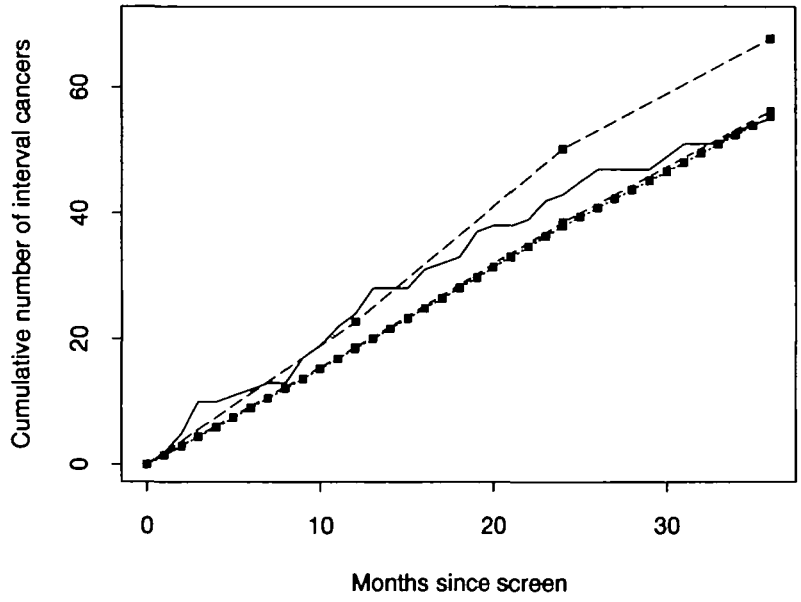

FIGURE 6. Observed (—) and maximum likelihood estimates of (model 1: – – –, model 2: ----, model 3: · · · ·) cumulative incidence of colorectal cancer following a negative screen at age 65–74 years, based on data from colorectal cancer screening in Calvados, France (1991–1994).

clinical disease was assumed to be equal to the incidence of cancers in the period just before screening. However, it would be more reasonable to allow for some variability in this estimate. We investigated the impact that this would have on the estimates of sensitivity and mean sojourn time, using the Gibbs sampling approach with a uniform prior distribution for preclinical incidence, J, over the range of 0.9–1.1 times the prescreening cancer incidence. The resulting estimates were little different from those obtained above. For example, model 2 applied to the data for 55- to 64-year-olds gave estimates for $\hat{\lambda}$ and S of 0.299 (95 percent credible interval 0.08–0.65) and 0.53 (95 percent credible interval 0.19–0.91), respectively.

Despite the reservations expressed above, the message from this modeling exercise is that the mean sojourn time is approximately 2 years at ages 45–54 years, increasing to about 6 years at ages 65–74. These mean sojourn time estimates reflect the relation between the variation in mean sojourn time according to subsite and the variation in mean age according to subsite (17). There should be some caution about this, because of the high correlation between sensitivity and mean sojourn time estimates, but results from randomized trials of Hemoccult screening indicate that the sensitivity estimates are roughly correct (10–12). This in turn suggests that the estimates of mean sojourn time are correct. For the purposes of understanding the disease's natural history and how it can be arrested by screening, targets for future research include refining the models and seeking explanations for age effects in terms of stage and for progression rates by subsite and histology.

REFERENCES

1. Stevenson CE. Statistical models for cancer screening. Stat Methods Med Res 1995;4:18–32.
2. van Oortmarssen GJ, Habbema JD, Lubbe JT, et al. A model-based analysis of the HIP project for breast cancer screening. Int J Cancer 1990;46:207–13.
3. Zelen M, Feinleib M. On the theory of screening for chronic disease. Biometrics 1969;56:601–14.
4. Prorok PC. The theory of periodic screening. II. Doubly bounded recurrence times and mean lead time and detection probability estimation. Adv Appl Probabil 1976;8:460–76.
5. Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. Biometrics 1984;40:1–14.
6. Prorok PC. Mathematical models and natural history in cervical cancer screening. In: Hakama M, Miller AB, Day NE, eds. Screening for cancer of the uterine cervix. (IARC Scientific Publication no. 76). Lyon, France: International Agency for Research on Cancer, 1986:185–96.
7. Albert A. Estimated cervical cancer disease state incidence and transition rates. J Natl Cancer Inst 1981;67:571–6.
8. van Oortmarssen GJ, Habbema JD. Epidemiological evidence for age-dependent regression of pre-invasive cervical cancer. Br J Cancer 1991;64:559–65.
9. Flehinger BJ, Kimmel M. The natural history of lung cancer in a periodically screened population. Biometrics 1987;43:127–44.
10. Eddy DM, Nugent FW, Eddy JF, et al. Screening for colorectal cancer in a high-risk population: results of a mathematical model. Gastroenterology 1987;92:682–92.
11. Swedish Cancer Society and the Swedish National Board of Health and Welfare. Breast-cancer screening with mammography in women aged 40–49 years. Int J Cancer 1996;68:693–9.
12. Walter SD, Day NE. Estimation of the duration of a pre-

clinical disease state using screening data. Am J Epidemiol 1983;118:865–86.
13. Paci E, Duffy SW. Modelling the analysis of breast cancer screening programmes: sensitivity, lead time and predictivevalue in the Florence District Programme (1975–1986). Int J Cancer 1991;20:852–8.
14. Gelfand AE, Smith AF. Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 1990;85: 398–409.
15. Gilks WR, Clayton DG, Spiegelhalter DJ, et al. Modelling complexity: applications of Gibbs sampling in medicine (with discussion). J R Stat Soc B 1993;55:39–102.
16. Thomas A, Spiegelhalter DJ, Gilks WR. BUGS: a program to perform Bayesian inference using Gibbs sampling. In: Bernando JM, Berger JO, Dawid AP, et al, eds. Bayesian statistics 4. Oxford, England: Clarendon Press, 1992: 837–42.
17. Launoy G, Smith TC, Duffy SW, et al. Colorectal cancer mass-screening: estimation of faecal occult blood test sensitivity, taking into account cancer mean sojourn time. Int J Cancer 1997;73:220–4.
18. Duffy SW, Chen HH, Tabar L, et al. 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. Stat Med 1995;14:1531–43.
19. Wolfram S. The Mathematica book. 3rd ed. New York, NY: Wolfram Media/Cambridge University Press, 1996.
20. SAS Institute, Inc. SAS/STAT user's guide. Version 6, vol 2. Cary, NC: SAS Institute, Inc, 1989.
21. Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences. Stat Sci 1992;7:457–72.

APPENDIX

BUGS Model Specification Code for Model 2

```
can.scr ~ dbin(p.scr,n.scr);
p.scr <- (S*J*(exp(-lam*age) - exp(-J*age))/
  (J - lam))/(exp(-J*age) + (J*(exp(-lam*age) -
  exp(-J*age))/(J - lam)));
for(i in 1:N){

        o[i] ~ dpois(theta.obs[i]);
  theta.new[i] <- py[i]*J*(1 - exp(-lam*
                  (time[i] - 0.5)));
  theta.und[i] <- can.scr*exp(logit.S)*
                  (exp(-lam*(time[i] - 1)) -
                  exp(-lam*time[i]));
  theta.obs[i] <- theta.new[i] + theta.und[i];
}

logit.S ~ dunif(-4,4);
S <- 1/(1 + exp(logit.S));
lam ~ dgamma(0.001,0.001)I(0.05,5);
mst <- 1/lam;
```