# Multi-state Markov models in cancer screening evaluation: a brief review and case study

Z. Uhry, G. Hédelin, M. Colonna, B. Asselain, P. Arveux, A. Rogel, C. Exbrayat, C. Guldenfels, I. Courtial, P. Soler-Michel, F. Molinié, D. Eilstein and SW Duffy

*Stat Methods Med Res* 2010 19: 463 originally published online 15 March 2010
DOI: 10.1177/0962280209359848

*Statistical Methods in Medical Research* 2010; **19**: 463–486

# Multi-state Markov models in cancer screening evaluation: a brief review and case study

**Z Uhry** Département des Maladies Chroniques et des Traumatismes, Institut de veille sanitaire, St-Maurice, France, **G Hédelin** Registre des cancers du Bas-Rhin, Laboratoire d'épidémiologie et de santé publique, EA3430, Université de Strasbourg, Strasbourg, France, **M Colonna** Registre des cancers de l'Isère, Meylan, France, **B Asselain** Service de biostatistique, Institut Curie, Paris, France, **P Arveux** Registre des cancers de Côte d'Or, Centre GF Leclerc, EA 4184, Dijon, France, **A Rogel** Département des Maladies Chroniques et des Traumatismes, Institut de veille sanitaire, St-Maurice, France, **C Exbrayat** Structure de gestion du dépistage de l'Isère, Meylan, France, **C Guldenfels** Structure de gestion du dépistage du Bas-Rhin, Strasbourg, France, **I Courtial, P Soler-Michel** Structure de gestion du dépistage du Rhône, Lyon, France, **F Molinié** Registre des cancers de Loire-Atlantique, Nantes, France, **D Eilstein** Département Santé Environnement, Institut de veille sanitaire, St-Maurice, France and **SW Duffy** Department of Epidemiology, Mathematics and Statistics, Cancer Research UK, London, United Kingdom

This work presents a brief overview of Markov models in cancer screening evaluation and focuses on two specific models. A three-state model was first proposed to estimate jointly the sensitivity of the screening procedure and the average duration in the preclinical phase, i.e. the period when the cancer is asymptomatic but detectable by screening. A five-state model, incorporating lymph node involvement as a prognostic factor, was later proposed combined with a survival analysis to predict the mortality reduction associated with screening. The strengths and limitations of these two models are illustrated using data from French breast cancer service screening programmes. The three-state model is a useful frame but parameter estimates should be interpreted with caution. They are highly correlated and depend heavily on the parametric assumptions of the model. Our results pointed out a serious limitation to the five-state model, due to implicit assumptions which are not always verified. Although it may still be useful, there is a need for more flexible models. Over-diagnosis is an important issue for both models and induces bias in parameter estimates. It can be addressed by adding a non-progressive state, but this may provide an uncertain estimation of over-diagnosis. When the primary goal is to avoid bias, rather than to estimate over-diagnosis, it may be more appropriate to correct for over-diagnosis assuming different levels in a sensitivity analysis. This would be particularly relevant in a perspective of mortality reduction estimation.

## 1 Introduction

The development of cancer screening trials and service screening programmes has led to a considerable methodological development in cancer screening evaluation in recent years.[1–8] The methods proposed involve modelling of the natural history of the disease and they can be roughly classified as likelihood- or simulation-based approaches.

Address for correspondence: Z Uhry, Département des Maladies Chroniques et des Traumatismes, Institut de veille sanitaire, 94 410 St-Maurice, France. E-mail: z.uhry@invs.sante.fr



10.1177/0962280209359848

Multi-state Markov models are common in the likelihood-based approaches,[9–26] although other models have also been used.[7,8,27–33] Simulation-based approaches were initially developed to carry out cost-effectiveness analysis. They rely on either micro-simulation[34–39] or on Markov decision-tree analysis.[40–42]

A simple three-state Markov model was first proposed to model the natural history of breast cancer.[9,10] The aim was to estimate jointly the sensitivity of the screening procedure and the mean sojourn time (MST), i.e. the average duration of the preclinical phase. The preclinical phase refers to the period when the cancer is asymptomatic but detectable by screening. These two parameters are of particular interest in screening evaluation. Multi-state models were also proposed to assess the more complex natural history of colorectal cancer.[19] Models involving prognostic factors for breast cancer were later proposed, coupled with survival analysis to predict the mortality reduction associated with screening.[12,13] The aim was to compare different screening frequencies[12,13] and to evaluate a service screening programme without a randomised control group.[16] More recently, models incorporating a non-progressive state have been proposed to evaluate over-diagnosis associated with screening.[24–26] Multi-state Markov models can address a variety of issues and have been used extensively. These models are simplistic, however, and rely on a number of assumptions, which should be examined carefully.

Our article focuses on the use of Markov models for two possible objectives in cancer screening evaluation: (1) to estimate the sensitivity of the screening procedure and the MST and (2) to predict the mortality reduction associated with screening, using the node status (i.e. whether the cancer has spread to the regional lymph nodes) as a prognostic factor. It aims to illustrate the strengths and limitations of multi-state Markov models with respect to these two objectives, using data from French breast cancer service screening programmes.

## 2 Material and methods

### 2.1 Modelling the natural history of breast cancer

The natural history of breast cancer, describing the progression of the tumour from the asymptomatic phase to the clinical phase, is modelled as multi-state regular homogeneous Markov models. Considering the natural history, together with the sensitivity of the screening procedure, allows us to model the screening process. The sensitivity of the screening procedure(s) is the probability that when the test is applied to a subject with the disease in the preclinical phase, a positive screening result is obtained.

Multi-state regular homogeneous Markov stochastic processes have well-defined properties:

(1) The transition rates between the states are invariant with time.
(2) The sojourn time in a given state follows an exponential distribution.
(3) The sojourn times in successive states are independent.
(4) The probability to transit from one state $a$, to another state $b$, is independent of the time spent in state $a$.
(5) The probability to transit from one state $a$, to another state $b$, is independent of the past states before $a$.

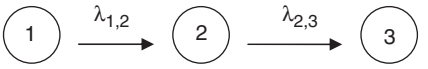


**Figure 1** Three-state model: State (1) no detectable breast cancer, State (2) Preclinical cancer, detectable by screening (asymptomatic), State (3) Clinical breast cancer (symptomatic). $\lambda_{1,2}$ and $\lambda_{2,3}$ are the transition rates between the states.

### *2.1.1 Three-state model*

The simplest model describes the progression of breast cancer through three successive states: no detectable cancer, preclinical cancer and clinical cancer. The preclinical phase is the period when the cancer is asymptomatic, without clinical signs, but detectable by screening mammography. When the cancer becomes symptomatic, it enters the clinical state and will be diagnosed without screening. Figure 1 illustrates this model. The MST is important in screening evaluation since it is related to the window of opportunity for early detection and to the potential advance in diagnosis. Work by Day and Walter based on data from the Health Insurance Plan supports the exponential distribution of sojourn time implicit in the Markov model, for breast cancer.[2,3] The exponential distribution implies that the average advance in diagnosis due to screening is equal to the MST and that the standard deviation of the sojourn time is equal to its mean.

The transition rate $\lambda_{1,2}$ represents the incidence rate of the preclinical disease. This rate can be assimilated to the usual incidence rate for a slightly older age due to the preclinical phase. The MST is equal to $1/\lambda_{2,3}$ under the exponential assumption.

The state at time $t$, $\{E_t\}$ is assumed to be a homogeneous Markov process, with instantaneous transition matrix $\Lambda$.

$$\Lambda = \begin{bmatrix} -\lambda_{1,2} & \lambda_{1,2} & 0 \\ 0 & -\lambda_{2,3} & \lambda_{2,3} \\ 0 & 0 & 0 \end{bmatrix}.$$

The matrix of finite transition probabilities $P_\lambda(t)$, which describes the probability of transiting from one state to another within the time $t$, can be derived as

$$P_\lambda(t) = \exp(\Lambda.t), \tag{1}$$

where $D$ is the diagonal matrix of the eigenvalues of $\Lambda$ and X is such that $\Lambda = X.D.X^{-1}$.

Formula (1) is a generic formula applicable to all homogeneous Markov processes. For the simple three-state model, the matrix $P_\lambda(t)$ can also be derived explicitly by integration of the appropriate convolutions of the exponential probability density functions.

$$P_\lambda(t) = \begin{bmatrix} e^{-\lambda_{1,2}t} & \frac{\lambda_{1,2}\left(e^{-\lambda_{1,2}.t} - e^{-\lambda_{2,3}.t}\right)}{\lambda_{2,3}-\lambda_{1,2}} & 1 - \frac{\left(\lambda_{2,3}.e^{-\lambda_{1,2}.t} - \lambda_{1,2}.e^{-\lambda_{2,3}.t}\right)}{\lambda_{2,3}-\lambda_{1,2}} \\ 0 & e^{-\lambda_{2,3}.t} & 1 - e^{-\lambda_{2,3}.t} \\ 0 & 0 & 1 \end{bmatrix}.$$

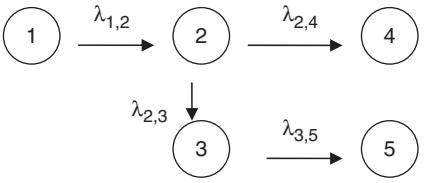


**Figure 2** Five-state model according to node status (negative = pN−/positive = pN+) State (1) no detectable breast cancer, State (2) Preclinical cancer pN−, detectable by screening (asymptomatic), State (3) Preclinical cancer pN+, detectable by screening (asymptomatic), State (4) Clinical breast cancer pN− (symptomatic), State (5) Clinical breast cancer pN+, (symptomatic). $\lambda_{1,2}$, $\lambda_{2,3}$, $\lambda_{2,4}$ and $\lambda_{3,5}$ are the transition rates between the states.

### *2.1.2 Five-state model according to node status*

The five-state model differentiates the progression of tumour according to node status. Incorporating a prognostic factor, coupled with survival analysis, one is able to predict mortality reduction associated with screening.[12,13,16] This model is presented in Figure 2.

As in the three-state model, the transition rate $\lambda_{1,2}$ represents the incidence rate of the preclinical disease. Sojourn times in the two preclinical states (node negative and node positive) both follow an exponential distribution (parameter $\lambda_{2,3} + \lambda_{2,4}$ and $\lambda_{3,5}$, respectively) and are assumed independent. As a consequence, MST in the preclinical phase for clinical node positive cancers (total sojourn time in states 2 and 3) is longer than for clinical node negative cancers (sojourn time in state 2 only).

As before, let $E_t$ be the state at time $t$. $\{E_t\}$ is assumed to be a homogeneous Markov process, with instantaneous transition matrix $\Lambda$.

$$\Lambda = \begin{bmatrix} -\lambda_{1,2} & \lambda_{1,2} & 0 & 0 & 0 \\ 0 & -(\lambda_{2,3} + \lambda_{2,4}) & \lambda_{2,3} & \lambda_{2,4} & 0 \\ 0 & 0 & -\lambda_{3,5} & 0 & \lambda_{3,5} \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

The probability matrix transition $P_\lambda(t)$ is calculated with formula (1), based on matrix calculations.

## 2.2 Likelihood and maximisation

In the context of a screening programme, tumours can arise by being detected at a round of screening (screen-detected), being diagnosed symptomatically between rounds of screening (interval cancer), or in subjects not attending for screening (non-attenders). Observed data consist in detection rates at different screening rounds and interval cancer rates, declined according to node status for the five-state model. These observations depend both on the natural history of the cancer and on the sensitivity of the mammography. In the five-state model, we assumed that the sensitivity of the mammography was not dependent on the node status. The likelihood is a function of the transition rates and sensitivity. Estimation of these parameters is realised by the maximisation of the log-likelihood. The likelihood derivation is straightforward when screening sensitivity

is 100%. It is more laborious otherwise, but it does not present theoretical difficulty. These models, which incorporate a sensitivity parameter, are a special case of multi-state Markov models with a classification error.[43]

Likelihood derivations for the three-state and the five-state models are presented in Appendix 1. The situation can be complicated by opportunistic screening outside the programme and unknown to the official programme provider. In this case, the likelihood can be modified to take into account assumptions on opportunistic screening. Regarding the five-state model analysis, cancers with unknown node status (pNx) must be incorporated in the likelihood. If available, auxiliary information on tumour size for pNx cancers may be exploited. In this case, the likelihood would include conditional probabilities of being node positive or node negative according to the tumour size.

These models assume constant transition rates over time. Theoretically, this assumes a constant preclinical incidence rate from birth to age at first screening, which is unrealistic. In practice, since all probabilities involved in the likelihood are conditioned on the fact that the woman did not have a clinical cancer before first screening (see Appendix), this cancels out the cumulative incidence before age at first screening.

Parameters are constrained, sensitivity of mammography is between 0 and 1, and transition rates are deemed to be positive. Programming used S-Plus7, with the function *nlminb* for box-constrained non-linear optimisation and *vcov.nlminb* for estimation of the variance–covariance matrix of the parameters. These functions are available in the library *mass*.[44] Confidence intervals for non-monotonic functions of parameters (e.g. MST and predictions in five-state model) were estimated empirically on 4000 simulations, assuming a multivariate normal distribution for parameter estimates and using their variance–covariance matrix.

## 2.3 Control group data

Markov models were initially proposed for randomised controlled trial data where control group data were available. Traditionally, in the three-state model, preclinical incidence was derived directly from control group data and entered as a known external parameter thus improving accuracy of the estimation of the other two parameters. In the five-state model analysis, control group data gave information both on incidence rate and node status distribution. This information could not be summarised through fixed values for some parameters. Control group data were then analysed jointly with screening data to estimate the parameters.[12,13]

However, estimation without control group data has been shown to be feasible for the three-state and five-state models, suggesting these models may be used to analyse service screening programmes.[11,16]

## 2.4 Illustration for the three-state model

Figure 3 presents expected cancer rates under the three-state model over a 8-year period, in the case of two screening rounds: detection rate at first screening, interval cancer rate after the first screening, detection rate at second screening and interval cancer rate after second screening, up to 8 years of follow-up after first screening. Several scenarios are presented: by the sensitivity of the screening procedure (100 or 80%), MST (2, 3 or 4 years) and screening interval (2 or 3 years). Preclinical incidence ($\lambda_{1,2}$)

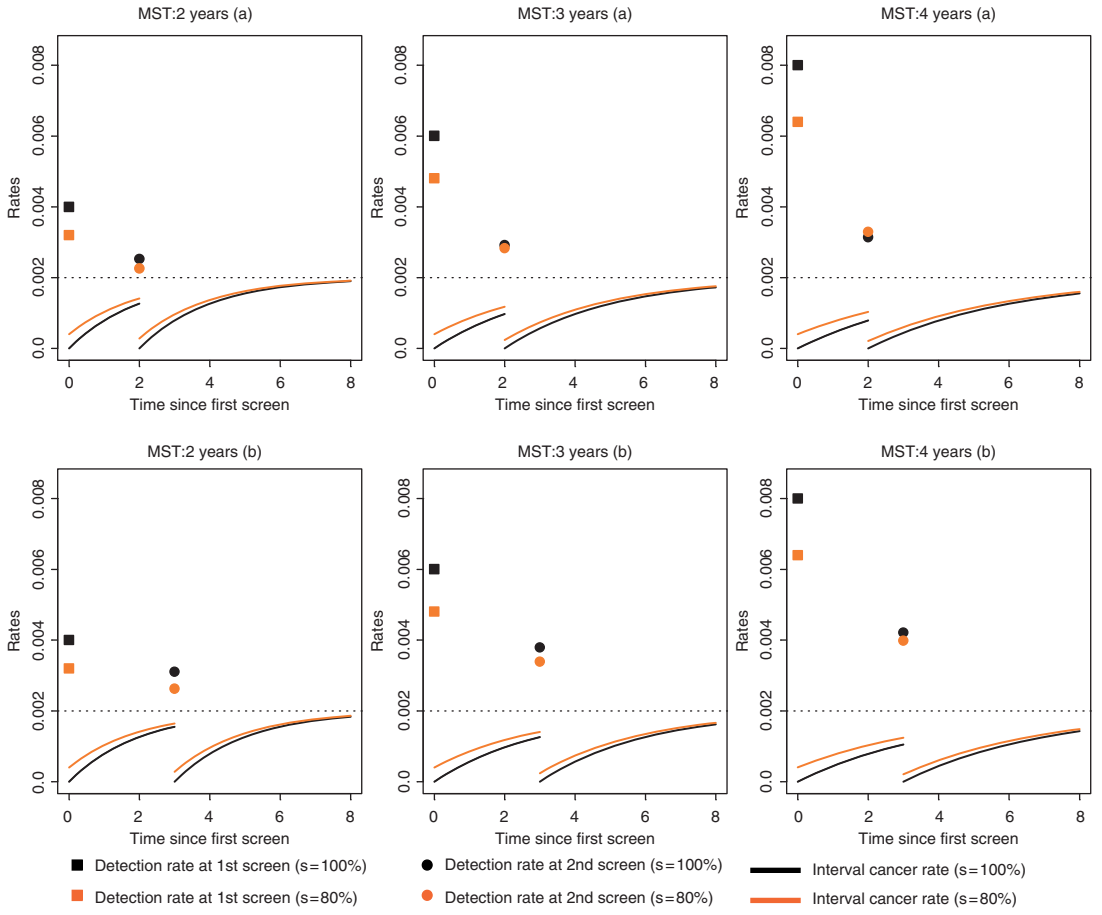


**Figure 3** Three-state model. Detection rates at first and second screenings and interval cancer rates for different MSTs and sensitivity, with preclinical incidence of 0.002: (a) 2-yearly screening; (b) 3-yearly screening.

was assumed to be 0.002 and is indicated by the dotted line. The detection rate at first screening increases linearly with increasing MST, as expected. This detection rate can roughly be approximated by $s.\lambda_{1,2}$ MST[6] (p. 129). It takes around twice the MST for interval cancers after the last screening to approach the expected incidence without screening. A lower screening sensitivity increases interval cancer rates, reduces screening detection rates, and reduces the detection rate ratio between the first and second screenings.

## 2.5 Different issues for different objectives

The three-state model focuses on estimating MST and sensitivity parameters per se. Conversely, the five-state model is a first step in predicting breast cancer mortality reduction associated with screening. It allows prediction of cancer cases by node status, in a situation with and without screening.[12,13,16] These predictions, coupled with survival

**Table 1** Description of the French data analysed (invasive breast cancers)

| | | Bas-Rhin | Isère | Rhône |
|---|---|---|---|---|
| Programme | Target age (years) | 50–64 | 50–69 | 50–69 |
| | Rhythm (years) | 2 | 2.5 | 3 |
| | Period analysed | 1989–1995 | 1990–1996 | 1987–1996[a] |
| Activity | Nb women | 52,071 | 37,381 | 93,532 |
| | Nb mammography | 107,523 | 54,339 | 141,288 |
| Screen-detected cancers | Cases (invasive) | 410 | 273 | 691 |
| | pNx (%) | 10 | 5 | 2 |
| Interval cancers | Cases (invasive) | 184 | 140 | 527 |
| | pNx (%) | 15 | 7 | 2 |
| Hypothesis on opportunistic screening[b] | Before first organised screening (%) | 35 | 55 | 55 |
| | Time interval[c] (years) | 2.5 | 2.5 | 2.5 |
| | Between organised screenings (%) | – | 15 | 25 |
| | Time interval[d] (years) | – | 1.5 | 1.75 |

[a]Pilot screening programme only part of the district before 1990.
[b]Proportions of women undergoing opportunistic screening among attendees to organised screening. Derived from self-reported mammography antecedents and from studies on detection mode of interval cancers. Opportunistic screening is assumed to have the same sensitivity as organised screening. We assumed one single previous opportunistic screening before first organised screening.
[c]Time in years between opportunistic screening before first screening and first screening.
[d]Time in years between organised screening and next opportunistic screening.

results by node status, give an estimate of the mortality reduction associated with screening. Here, the focus is on mortality prediction rather than on the natural history and screening parameters themselves.

## 2.6 Data for case-study

Data are from breast screening programmes in three districts in France: Bas-Rhin, Isère and Rhône. Pilot programmes of organised screening were launched in France around 1990 in ten districts. A nationally organised screening programme was adopted in 1994 and reached full coverage in 2004. Opportunistic screening outside the programme was widespread. The pilot programmes targeted women aged 50–69 years and the frequency of invitation varied between 2 and 3 years.

We studied an early period (1990–1995) during the pilot programmes because a long follow-up for survival analysis was necessary to realise an estimation of the mortality reduction associated with screening. Over this period and among districts involved in the pilot programmes only two were covered by a cancer registry (Bas-Rhin and Isère). Another district (Rhône), however, had a system of multi-source ascertainment of breast cancer. Basic population and screening data for these three districts are given in Table 1. Interval cancers are defined as cancers arising after a negative screening and before next invitation, whatever be the real detection mode of the cancer.

Opportunistic screening before entering the organised programme was common and also existed, to a lesser extent, between screenings. We made assumptions about opportunistic screening before entering the programme and between two organised screenings

(Table 1) based on self-reported antecedent mammography, collected by local management centres, and on two specific studies conducted in Rhône and Isère on the detection mode of interval cancers. Opportunistic screening is assumed to have the same sensitivity as organised screening. We assumed one single, previous, opportunistic screen before the first organised screening.

In the French screening context, there is no suitably comparable unscreened population. However, one cancer registry (Loire-Atlantique) collected the detection mode of cancers in our study period (clinical cancer or screen-detected by opportunistic screening). Organised screening started later in this district. This provided information on the distribution of node status among clinical unscreened cancers.

### 2.7 Analyses

For the three-state model data were first analysed separately for the three districts. Thereafter, data were analysed jointly, assuming separate preclinical incidence rates for each district, but common transition rates once the cancer entered the preclinical phase, and common sensitivity. Sensitivity of the mammography may be tested for variation among districts using the usual likelihood ratio test. In the five-state model, two analyses were carried out, with and without data from clinical cancers in Loire-Atlantique, to represent an unscreened population.

Overall chi-squared measures of goodness of fit were calculated, comparing observed numbers of cases to estimated numbers, by detection mode (cancers detected at first screening, cancers detected at subsequent screening and interval cancers) and district. We present summary measures of goodness of fit to indicate the main elements of discussion for each model.

## 3 Results

### 3.1 Three-state model

Table 2 presents the analyses for each district separately. MST in the preclinical phase varied from 2.3 to 3.3 years. Sensitivity varied from 86% to 94%. Chi-squared goodness of fit was excellent in Bas-Rhin and poorer in Isère and Rhône, due to lack of fit for interval cancer rates. In Isère, the number of interval cancers was over-estimated 12–23 months after the last screening and under-estimated 24–29 months after the last screening, and conversely in Rhône. This might be due to more complex patterns of opportunistic screening than we assumed. Parameter estimates were highly correlated and different parameter values could give close predicted values.

Table 3 presents the multi-district analyses, by age group. Districts were analysed jointly, assuming different preclinical incidence rates but common sensitivity and MST. All parameter estimates showed consistent variation with age, with increasing preclinical incidence rates, sensitivity of the mammography and MST, as expected.[12] Overall goodness of fit was reasonable for the age group 50–59 years and poorer for 60–69 years and for all ages. Lack of fit essentially concerns interval cancers in Isère and Rhône.

Table 4 presents a sensitivity analysis related to opportunistic screening. The level of opportunistic screening, between screenings and before first screening separately, increased or decreased, respectively, in all districts at the same time. Parameter estimates

**Table 2** Three-state model. Analyses by district (50–69 years)

| District | $\lambda_{1,2}$[a] | $\lambda_{2,3}$ | s | MST[b] [CI 95%] | Chi[c] |
|---|---|---|---|---|---|
| Bas-Rhin (BR) | 2.4 | 0.43 | 0.94 | 2.3 | 0,2 |
| (SE)[bb] | (0.1) | (0.06) | (0.07) | [1.8–3.4] | |
| Isère (I) | 2.6 | 0.34 | 0.92 | 2.9 | 5,8 |
| (SE)[b] | (0.2) | (0.06) | (0.08) | [2.2–4.3] | |
| Rhône (R) | 2.5 | 0.31 | 0.86 | 3.3 | 7,8 |
| (SE)[b] | (0.1) | (0.04) | (0.06) | [2.6–4.3] | |

[a] × 1000.
[b] MST in years.
[c] Chi-squared goodness of fit; df = 1 for Bas-Rhin (threshold 3.8) and df = 2 for Isère and Rhône (threshold 6.2).
[d] Standard error.

**Table 3** Three-state model. Multi-districts analyses, by age group

| Age | $\lambda_{1,2}^{BR}$[a] | $\lambda_{1,2}^{I}$[a] | $\lambda_{1,2}^{R}$[a] | $\lambda_{2,3}$ | s[b] | MST[c] [CI 95%] | Chi[d] (df = 9) |
|---|---|---|---|---|---|---|---|
| 50–59 years | 2.2 | 2.3 | 2.4 | 0.42 | 0.87 | 2.4 | 11.6 |
| (SE)[e] | (0.1) | (0.2) | (0.1) | (0.05) | (0.06) | [1.9–3.2] | |
| 60–69 years | 2.4 | 2.9 | 2.6 | 0.30 | 0.95 | 3.3 | 18.6 |
| (SE)[e] | (0.2) | (0.2) | (0.1) | (0.03) | (0.04) | [2.8–4.0] | |
| 50–69 years | 2.2 | 2.6 | 2.6 | 0.35 | 0.90 | 2.8 | 19.3 |
| (SE)[e] | (0.1) | (0.1) | (0.1) | (0.03) | (0.04) | [2.4–3.3] | |

[a] × 1000 (BR: Bas-Rhin, I: Isère, R: Rhône).
[b] Test of equality of the sensitivity in the 3 districts: $p = 0.35$, $p = 0.92$ and $p = 0.09$, respectively, in the different age groups (50–59, 60–69 and 50–69 years).
[c] MST in years.
[d] Chi-squared goodness of fit; threshold 16,9.
[e] Standard error.

were not sensitive to hypotheses on opportunistic screening before the first screening but were sensitive to hypotheses on opportunistic screening between screenings, in particular the MST.

We carried out a simulation study to explore the effect of over-diagnosis on parameter estimates. In the context of cancer screening, over-diagnosis refers to the detection of cancer by screening which would never have become symptomatic in a woman's lifetime, either because of death from competing risks during the preclinical phase or because the tumour is non-progressive. Because breast cancer screening targets middle-aged women in developing countries, who tend to have a long future-life expectancy, over-diagnosis is mainly related to detection of non-progressive tumours. The simulations were derived by adding a non-progressive preclinical state to the three-state model (state four, transition rate $\lambda_{1,4}$ from the non-disease state). This state represents tumours detectable by screening that would never arise clinically in the absence of screening. This four-state model was used to simulate data, which were then analysed assuming the three-state model. Parameters estimates were then compared with those used for simulation (Table 5). Sensitivity estimation was largely unaffected, whereas preclinical

**Table 4** Three-state model, multi-districts analysis. Sensitivity analysis to opportunistic screening hypotheses (50–69 years)

| Hypothesis | $\lambda_{1,2}^{BR}$[a] | $\lambda_{1,2}^{I}$[a] | $\lambda_{1,2}^{R}$[a] | $\lambda_{2,3}$ | s | MST[b] |
|---|---|---|---|---|---|---|
| Original | 2.2 | 2.6 | 2.6 | 0.35 | 0.90 | 2.8 |
| Between screenings, ↓ 10%[c] | 2.3 | 2.7 | 2.6 | 0.37 | 0.91 | 2.7 |
| Between screenings, ↑ 10%[d] | 2.1 | 2.5 | 2.4 | 0.31 | 0.87 | 3.2 |
| Before first screenings, ↓ 10%[c] | 2.2 | 2.6 | 2.5 | 0.37 | 0.91 | 2.7 |
| Before first screenings, ↑ 10%[d] | 2.2 | 2.7 | 2.6 | 0.34 | 0.90 | 2.9 |

[a] ×1000 (A: Bas-Rhin, B: Isère, C: Rhône).
[b] MST in years.
[c] Hypothesis about proportion of women undergoing opportunistic screening decreased by 10% in all districts.
[d] Hypothesis about proportion of women undergoing opportunistic screening increased by 10% in all districts.

**Table 5** Three-state model. Bias in parameter estimations induced by over-diagnosis (simulations)

| Parameters | $\lambda_{1,4}$[a] | $\lambda_{1,2}$[a] | $\lambda_{2,3}$ | s | MST[b] |
|---|---|---|---|---|---|
| Assumed[c] | 0.17[e] | 2.0 | 0.33 | 1.00 | 3.0 |
| Estimated[d] | – | 2.2 | 0.27 | 0.99 | 3.7 |
| Assumed[c] | 0.17[f] | 2.0 | 0.33 | 0.85 | 3.0 |
| Estimated[d] | – | 2.2 | 0.28 | 0.86 | 3.6 |
| Assumed[c] | 0.38[g] | 2.0 | 0.33 | 1.00 | 3.0 |
| Estimated[d] | – | 2.5 | 0.24 | 0.99 | 4.2 |
| Assumed[c] | 0.38[h] | 2.0 | 0.33 | 0.85 | 3.0 |
| Estimated[d] | – | 2.5 | 0.24 | 0.87 | 4.6 |

[a] ×1000.
[b] MST corresponds to the mean sojourn time in the progressive preclinical state, in years.
[c] Parameters of the four-state model used to simulate samples, adding a non-progressive state to represent over-diagnosis.
[d] Average parameter estimations obtained with the three-state model on 400 simulations of samples of 200,000 women undergoing two screenings at 2- and 5-year intervals.
[e] Corresponds to 23% over-diagnosed cancers at first screening and 11% at second screening.
[f] Corresponds to 23% over-diagnosed cancers at first screening and 16% at second screening.
[g] Corresponds to 40% over-diagnosed cancers at first screening and 22% at second screening.
[h] Corresponds to 40% over-diagnosed cancers at first screening and 29% at second screening.

incidence and MST were over-estimated when ignoring over-diagnosis. Preclinical incidence and MST refers, in this case, to the *progressive* preclinical state. Note that 20% of cancers over-diagnosed at the first screening and 10% at the second screening had already induced a significant bias. Constraining incidence from a control group as a fixed value for $\lambda_{1,2}$ would increase over-estimation of the MST and lead to under-estimation of the sensitivity.

## 3.2 Five-state model

For brevity, we present only results for all age groups combined. Table 6 presents the distribution of node status, according to detection mode. The proportion of node

**Table 6** Data description. Node status according to detection mode (invasive cancers)

| District | Cancers at first screening | | | Cancers at subsequent screenings | | | Interval cancers | | | Clinical cancers | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | N[a] | pN−[b] | pNx[c] | N[a] | pN−[b] | pNx[c] | N[a] | pN−[b] | pNx[c] | N[a] | pN−[b] | pNx[c] |
| Bas-Rhin | 216 | 74% | 10% | 154 | 69% | 9% | 157 | 59% | 15% | | | |
| Isère | 206 | 76% | 4% | 54 | 63% | 8% | 130 | 71% | 7% | | | |
| Rhône | 491 | 71% | 2% | 185 | 72% | 2% | 514 | 64% | 2% | | | |
| Loire-Atlantique[d] | | | | | | | | | | 825 | 52% | 1% |

[a]Number of cancers with known node status.
[b]Proportion of node negative cancers among known node status.
[c]Proportion of cancers with unknown node status.
[d]Women aged 50–69 years, diagnosed from 1991–1995; Recoding of the histological node status by the clinical node status for cancers without lymph node dissection and for cancers node negative after neo-adjuvant chemotherapy.

**Table 7** Five-state model, multi-districts analysis (50–69 years). Parameters estimates

| | $\lambda_{1,2}^{BR}$[a] | $\lambda_{1,2}^{I}$[a] | $\lambda_{1,2}^{R}$[a] | $\lambda_{2,3}$ | $\lambda_{2,4}$ | $\lambda_{3,5}$ | s | pN−[b] | MST pN+[b] | All[c] | Chi[d] (df = 11) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| LA not included[e] | 2.3 | 2.7 | 2.6 | 0.29 | 0.24 | 0.60 | 0.86 | 1.9 | 1.7 | 2.8 | 32.0 |
| (SE)[f] | (0.1) | (0.2) | (0.1) | (0.03) | (0.02) | (0.06) | (0.03) | | | | |
| LA included[e] | 2.2 | 2.6 | 2.6 | 0.25 | 0.25 | 0.54 | 0.83 | 2.0 | 1.9 | 2.9 | 38.1 |
| (SE)[f] | (0.1) | (0.2) | (0.1) | (0.03) | (0.02) | (0.06) | (0.04) | | | | |

[a]×1000 (BR: Bas-Rhin, I: Isère, R: Rhône).
[b]MST in preclinical node negative state (state 2; MST= $1/(\lambda_{2,3}+\lambda_{2,4})$) and preclinical node positive state, respectively (state 3, MST= $1/\lambda_{3,5}$).
[c]Overall MST in preclinical phase; MST = $\frac{1}{\lambda_{2,3}+\lambda_{2,4}} + \frac{1}{\lambda_{3,5}} \times \frac{\lambda_{2,3}}{\lambda_{2,3}+\lambda_{2,4}}$ (MST in state 2 + MST in state 3 × proportion of cancers moving to state 3 from state 2).
[d]Chi-squared goodness of fit; threshold 19.7. Fit concerns only the districts of Bas-Rhin, Isère, Rhône and even when Loire-Atlantique data are included.
[e]Loire-Atlantique (LA).
[f]Standard error.

negative cancers was higher at first screening compared to subsequent screenings in Bas-Rhin and Isère. This is unexpected in screening results. Table 6 also shows the proportion of node negative tumours observed in clinical cancers from Loire-Atlantique. Although not a proper control group, we may presume that this proportion represents a lower boundary for the situation without screening.

Table 7 shows the estimates in the five-state model multi-district analysis. The chi-squared statistics measures the overall goodness of fit, by comparing observed to predicted numbers of cases, by node status and detection mode. Without the Loire-Atlantique data, MST and preclinical incidence rates were close to those from the three-state model, while the sensitivity estimate was a little lower. Overall fit was poor. The lack of fit mainly concerns the cancers detected at subsequent screenings in Bas-Rhin and Isère, with over-estimation of proportion of node negative cancers. It also

**Table 8** Five-state model, multi-districts analysis (50–69 years). Predicted proportion of node negative cancers, with and without screening[a]

| | LA not included | | LA included | |
|---|---|---|---|---|
| | %pN− pred | CI 95% | %pN− pred | CI 95% |
| Cancers at first screening | 68 | [64–70] | 68 | [65–71] |
| Cancers at second/third screening | 78 | [76–80] | 79 | [76–80] |
| Interval cancers after first screening | 61 | [57–65] | 64 | [60–67] |
| Interval cancers after second/third screening | 66 | [62–70] | 70 | [67–73] |
| Screened population (overall) | 71 | [69–73] | 72 | [70–74] |
| Unscreened population | 45 | [42–49] | 49 | [47–52] |

[a]Over 6 years, assuming three screenings with a 2-year interval between screenings.

concerned node positive interval cancers, which were under-estimated in Bas-Rhin and over-estimated in Isère. Including Loire-Atlantique mainly affects estimates of sensitivity and of certain transition rates ($\lambda_{2,3}$ and $\lambda_{3,5}$) and worsens the fit.

Table 8 presents the model's prediction of the proportion of node negative tumours, in a situation with and without screening. Without using the Loire-Atlantique data, prediction of proportion of node negative cancers without screening is 45%. This is low compared to the proportion observed in clinical cancers in Loire-Atlantique (52%). However, including the Loire-Atlantique data did not solve the problem and led to an increased estimated proportion of node negative tumours among interval cancers. These results reveal a limitation of the model. When the proportion of node negative tumours among intervals cancers is not high enough compared to an unscreened population, the model cannot predict both proportions correctly.

As an illustration, in the Swedish Two County trial data, the proportion of node negative tumours was almost equal in interval cancers and in the control group for women aged 50–59 years or 60–69 years at randomisation (53% and 59% respectively).[12,13] Table 9 presents analyses for these age groups in the Swedish Two County trial data. The limitation is clear on these data. With or without control group data, the model does not reliably predict the situation without screening. It leads to over-estimation of the benefit of screening in terms of node status. Interestingly, this limitation was not observed in the 40–49 years age group.[16]

Sensitivity analyses concerning opportunistic screening and over-diagnosis were also carried out for the five-state model (data not shown). With respect to our objective, it is relevant to study these issues directly on the mortality reduction estimation, obtained by combining results from the five-state model with those of survival analysis. In particular, over-diagnosis also impacts on screen-detected cancer survival. This is beyond the scope of the present article. However, we can briefly summarise the main results of the sensitivity analyses for the five-state model. They were consistent with those for the three-state model. Parameter estimates and predicted proportions of node negative tumours were not sensitive to variations in the level of opportunistic screening before the first screening. They were a little sensitive to variations in the level of opportunistic screening between organised screenings. Simulations indicated that over-diagnosis induced an

**Table 9** Five-state model, Swedish Two County trial data. Predicted proportion of node negative cancers, with and without screening[a]

| | 50–59 years[b] | | 60–69 years[b] | |
|---|---|---|---|---|
| | Without control group | With control group | Without control group | With control group |
| Cancers at first screening | 77 | 78 | 82 | 81 |
| Cancers at second/third screening | 81 | 84 | 86 | 87 |
| Interval cancers after first screening | 50 | 61 | 53 | 64 |
| Interval cancers after second/third screening | 51 | 63 | 58 | 70 |
| Screened population (overall) | 71 | 75 | 78 | 80 |
| Unscreened population | 37 | 47 | 43 | 54 |

[a]Over three screens with a 33-months interval between screen, which corresponds to a 8.25 years period.
[b]Age at randomisation (while we used age at the screening mammography in the analysis of the French data).

over-estimation of preclinical incidence and of MST in the preclinical node negative state.

## 4 Discussion

The case-study illustrates some issues regarding the use of multi-state Markov models for two possible objectives: estimating sensitivity and MST, and predicting the node status of tumours diagnosed, which could in turn be used to estimate the mortality reduction associated with screening. The French context is characterised by a high level of opportunistic screening. We had to make simplistic assumptions about this practice. This adds uncertainty to our results.

Overall, we obtained plausible estimations with the three-state model, consistent with the literature.[3,6,9,15,18,45,46] The fit was reasonable given the simplistic assumptions adopted about opportunistic screening. Parameter estimates exhibited the anticipated variation with age. Incidence estimation in the multi-district analysis was consistent with cancer registry data of Bas-Rhin and Isère, lying between observed incidence before organised screening (1985–1988) and during the study period (1990–1995). Results, however, were sensitive to assumptions regarding opportunistic screening between screenings.

The three-state model was the first global framework used to estimate MST and sensitivity simultaneously. It is relatively easy to implement with available statistical software. It only requires matrix calculations abilities and optimisation procedures. Bayesian estimation is another possibility.[20,23] Initial work by Day and Walter supported the exponential distribution for the sojourn time in the preclinical phase for breast cancer.[2,3] The three-state model usually gives plausible parameter estimates, showing the anticipated variation with age, as observed in our case study. The model has been used to compare data from different countries and different screening programmes.[11]

Different datasets may be analysed jointly, with both common and dataset-specific parameters. The likelihood can be modified to take into account opportunistic screening among attendees at organised screenings. Three-state models with more flexible distributions for the sojourn time have also been proposed.[29–31]

This simple model, however, has several drawbacks. The first is a generic limitation to any model attempting to estimate jointly sensitivity and MST. Without information from a control group data on incidence, the estimates are highly correlated. This may lead to large confidence intervals and the model can produce similar predicted values with quite different parameters. This also makes the estimates sensitive to small changes in interval cancer data. This is less marked when preclinical incidence is constrained. Departures from the exponential distribution will also affect the estimates.[29–31] In particular, estimation of the sensitivity must be interpreted with caution. The preclinical phase and screening sensitivity in Markov models refers to cancers 'detectable' by screening. The models cannot infer with respect to the development of tumours before screening detectability, since it is not observable. This might, however, limit the comparability of results of screening tools with different detection abilities.[21,22,27] Estimations of screening sensitivity and MST should be considered jointly when interpreting and comparing different screening programme results.[21,22,27] These two parameters may indeed compensate for each other, yet provide reasonable predicted values. As an illustration, Weedon-Fekjaer *et al.* obtained a long MST and low-sensitivity analysing data from the Norwegian screening programme.[21,22]

Another drawback concerns bias induced by potential over-diagnosis, which leads to over-estimation of the MST. This is exacerbated, and sensitivity can be biased, if control group data are used to infer preclinical incidence. In the past, a four-state model including a non-progressive state, has been used to evaluate over-diagnosis from the data simultaneously with other parameters.[25,26] Evaluating over-diagnosis is of high interest and obviously more satisfactory than a simple bias simulation study. Parameter estimates can, however, be imprecise in the four-state model, with some exceptions.[25,26] Correlation of the parameter estimates is even stronger than in the three-state model. Given the uncertainty surrounding opportunistic screening in our data, an accurate estimation of the four-state model was considered beyond our objective. *In situ* cancers raise a similar issue. Some *in situ* cancers do not become invasive if undetected and untreated. The three-state model does not include this source of heterogeneity. The results of two analyses, *in situ* excluded or included, are often presented to give a range of estimates.[6] The above analyses included invasive cancers only. We obtained slightly higher estimates of MST (3.0) and sensitivity (92%) when including *in situ* cancers in the analysis, and higher estimation of preclinical incidence. A study proposed a six-state model integrating additional states to represent non-progressive and progressive *in situ* cancers.[24] Although sensitivity of the mammography had to be assumed to be 100% in this model, it explicitly modelled progression and non-progression of *in situ* cancers and provided estimates of all other relevant parameters. Models including additional states to represent non-progressive cancers, whether invasive or *in situ*, would be attractive but would suffer from a high correlation between parameter estimates.

The five-state model is a little more complex than the three-state model. Once the likelihood is derived, however, it is relatively easy to implement with available statistical software. Taking into account opportunistic screening is possible though laborious.

The five-state model has a different objective. It is used, coupled with survival analysis results, to predict the mortality reduction associated with screening. Our case study concerned the estimation of the five-state model parameters, but not the further combination with survival results. The model aims to predict breast cancer cases by node status in the situation with and without screening. Focus is on prediction of numbers of node positive and node negative cancers rather than the parameters *per se*.

Parameter estimates in the five-state model were consistent with those in the three-state model, although sensitivity of the mammography was a little lower. Prediction of the proportion of node negative tumours at second and subsequent screens was over-estimated. The lack of fit essentially concerns a configuration specific to our data in that the observed proportion of node negative tumours was smaller at subsequent screenings compared to the first screening in two districts. Actually, the model is unable to predict such a configuration. It provides predictions that conform more to expected screening results, suggesting the five-state model is adapted, on this point, to screening process modelling.

However, the validity of some assumptions in the five-state model may be questioned. In particular, sojourn times in the two preclinical states (node negative and node positive) were assumed independent, while they might be positively correlated. Another consequence of the model assumptions is that MST is assumed longer for clinical node positive than for clinical node negative cancers. Because interval cancers have shorter sojourn times on average than overall cancers, the model thus implies that interval cancers are more frequently node negative than unscreened cancers. This explains why the model cannot fit, if the proportion of node negative tumours in interval cancers is not high enough compared to an unscreened population. The implicit assumption of a longer MST for node positive compared to node negative clinical tumours might not hold, at least for a subgroup of fast growing aggressive cancers. Our results indicate that this assumption might not be verified in our data, nor in the Swedish Two County trial data for women aged 50–59 years or 60–69 years, while it seems more adapted for women aged 40–49 years.[16]

This is a serious limitation of the five-state model, which hasn't been noted previously.[12,13,16] These models should not be used to predict an unscreened population without data describing this situation. The predicted proportion of node negative tumours without screening may be under-estimated. On the other hand, including such data is a poor solution since the benefit of screening is still over-estimated in terms of node status distribution. A second best solution would be to use the model, based only on screening data, to predict the situation with screening and to use data from an unscreened population independently to predict the situation without screening. This would lead to the best possible fit on screening data and would avoid a bias in favour of screening. A better solution, obviously, would be to adopt more flexible models. In our case-study, using Loire-Atlantique data, the problem is apparent but not major. However, the problem could be more crucial in other cases.

In the model, we assumed that sensitivity of the mammography was the same in the preclinical phase, whatever the node status. This is probably untrue, since node status is correlated with size and sensitivity increases with tumour size. Sensitivity is certainly higher among node positive tumours. Incorporating two parameters for sensitivity, however, would represent a serious complication to the model. An accessible

alternative would be to assume 100% sensitivity for node positive tumours. This might not be true either, but it maximises the difference in sensitivities according to node status, representing the 'opposite' assumption compared to the main model. We tested this alternative. It hardly improved the fit; MST estimates were affected a little but the predictions were almost unchanged. Overall, this alternative presented the same limitations and inadequacies as observed in the main analysis. These results underline that assuming a common sensitivity whatever the node status is not a critical issue and does not explain the limitations that we have pointed out. In the five-state model analysis, the test of equality of the sensitivities between districts provided slightly different results than those obtained in the three-state model analysis ($p = 0.15$, $p = 0.38$ and $p = 0.01$, respectively, in the different age groups of 50–59, 60–69 and 50–69 years). The stronger evidence for heterogeneity of sensitivity in all-ages analysis is probably exacerbated by the difference in age distribution between districts. As mentioned, age-stratified analysis is more relevant but was not presented for brevity. Chi-squared measures of goodness of fit are around 22 in both age-stratified analyses. In all-ages analysis, allowing for district-specific sensitivities however didn't affect the transition rates estimates but improved the fit, mainly in Bas-Rhin. Allowing district-specific transition rates in the model did not further improve the fit. The limitation of the five-state model that we have pointed out was not related to this issue and also concerned single-districts as well as age-stratified analyses.

Varying opportunistic screening assumptions only slightly influenced the overall difference in node status distribution between the screening and non-screening situations. The mortality reduction estimation is therefore unlikely to be strongly affected. Simulation indicated that over-diagnosis induced over-estimation of preclinical incidence and of MST in the preclinical node negative state. Coupled with bias in survival of screen-detected cancers, this would probably induce a bias in favour of screening in the mortality reduction estimation. This could be handled theoretically by adding a non-progressive state to the model. Our data were too strongly affected by opportunistic screening to make such an inference. Furthermore, our primary goal is to avoid bias rather than to estimate over-diagnosis. It may then be more appropriate to correct for over-diagnosis, assuming various levels in a sensitivity analysis. This would avoid relying on estimates of over-diagnosis of uncertain validity and considerable imprecision. This is particularly relevant in a perspective of mortality reduction estimation. Correction and sensitivity analysis should be studied in the mortality reduction estimation itself.

These models were first proposed by Chen and Duffy to compare different screening rhythms (1, 2 or 3 years) in different age groups.[12,13] Several multi-state models were studied, according to size, node status, grade or a combination of these characteristics on the data from the Swedish Two County trial. Later, the authors used this approach to analyse a service screening programme without control group data.[16] The multi-state model was used to predict breast cancer cases over a study period according to tumour characteristics with and without screening. Survival results, derived from a multivariate analysis, were used to translate the shift in the tumours characteristic due to screening to an estimated effect on mortality. Chen *et al.* used the same term (10 years) to predict deaths for all cases and used the same survival results whatever the detection mode. An alternative would be to predict deaths at a fixed date after the start of the study period using screen-detected cancer survival for these cancers. This alternative attempts

to model mortality analogous to a randomised trial, at a fixed date after the start of the trial. Cases should be predicted in this case yearly and the term for predicting death would depend on the year of diagnosis. Using screen-detected cancer survival is relevant since the advance in diagnosis is handled by the multi-state model through the year of diagnosis. On the other hand, this solution would also require adjustment for selection bias and length bias.

Progression models are essential in cancer screening evaluation and there is a need for more flexible models. Semi-Markov models represent one possibility.[47–50] Foucher and colleagues proposed semi-Markov multi-state models for chronic diseases,[47] with application to renal transplantation,[48] HIV infection[49] and colorectal cancer control after curative resection.[50] The semi-Markov processes allow transitions out of a state to depend on the duration spent in this state. The sojourn time in a state follows a generalised Weibull distribution, corresponding to a large spectrum of hazard rates. In addition, the approach implemented allowed including covariates affecting the transition rates.[47–50] This general framework introduces a considerable flexibility compared to the Markov model we presented. To our knowledge, these semi-Markov models haven't been applied to cancer screening evaluation. Note that incorporating screening sensitivity in these models is not straightforward.

Another framework has also been proposed, based on biological tumour growth models,[27] breaking with multi-state modelling. Weedon-Fekjaer *et al*. implemented a tumour growth model of breast cancer combined with a continuous function of screening sensitivity according to tumour size.[27] This approach is based on a likelihood maximisation. It is particularly parsimonious, with only four parameters involved, thanks to the continuous modelling of both tumour growth and screening sensitivity. A specific variance parameter accounts for individual variation in tumour growth (random effect). In this approach, sojourn time in the preclinical phase does not follow a well-defined parametric distribution. In the case-study of the Norwegian screening data, the estimated individual variation in sojourn time was larger than that implicit in the exponential distribution. This suggests that the simple Markov model may assume too little individual variation in cancer progression. Note that implementation of this approach is not easy and implies mixed-effects estimation for tumour growth. Modelling the underlying biological processes is undoubtedly an interesting perspective.[8,27,39]

## 5 Conclusion

### 5.1 Three-state model: useful frame, cautious interpretation

.

The three-state model is a useful frame and the exponential assumption was supported by initial analyses of breast cancer data. It is easy to implement and usually provides plausible estimates. However, estimates are highly correlated and depend heavily on the parametric assumptions of the model. In particular, estimation of the sensitivity should be considered cautiously and interpreted together with the MST estimate. This is a generic limitation when attempting to estimate MST and sensitivity of the screening procedure jointly.

### 5.2 Models incorporating prognostic factors: need for more flexible models

Our results pointed out a serious limitation to the five-state Markov model, though it may still be useful. Some assumptions are not appropriate for certain datasets. The French data pose particular challenges, because of the absence of suitable unscreened comparison groups and the uncertainties induced by opportunistic screening. However, the limitation of the model was even clearer using data from a randomised trial. Progression models are essential in cancer screening evaluation and there is a need for more flexible models.

### 5.3 Over-diagnosis: correction and sensitivity analyses

Over-diagnosis is an important issue, especially in the mortality reduction estimation. This issue can be addressed in a global attempt of handling and estimating over-diagnosis adding a non-progressive state to the model. These models are attractive, but estimation of over-diagnosis may be uncertain. When the primary goal is to avoid biased estimations rather than to estimate over-diagnosis, it may be more appropriate to correct for over-diagnosis, assuming different levels in a sensitivity analysis. It would be particularly useful to carry out the sensitivity analysis on the mortality reduction estimation itself.

## References


1 Prorok PC. Mathematical models for breast cancer screening. In Day NE, Miller AB, eds. *Screening for breast cancer*. Hans Huber, Toronto; 1988: 95–109.

2 Walter SD, Day NE. Estimation of the duration of a preclinical disease state using screening data. *American Journal of Epidemiology* 1983; **118**: 865–86.

3 Day NE, Walter SD. Simplified models of screening for chronic disease: estimation procedures from mass screening programmes. *Biometrics* 1984; **40**: 1–14.

4 Stevenson CE. Statistical model for cancer screening. *Statistical Methods in Medical Research* 1995; **4**: 18–32.

5 Van Oortmarssen GJ, Boer R, Habbema JD. Modelling issues in cancer screening. *Statistical Methods in Medical Research* 1995; **4**: 33–54.

6 Duffy SW, Hill C, Estève J. *Quantitative methods for the evaluation of cancer screening*. Oxford University Press, New York; 2001.

7 Boer R, Plevritis S, Clarke L. Diversity of model approaches for breast cancer screening: a review of model assumptions by the Cancer Intervention and Surveillance Network (CISNET) Breast Cancer Groups. *Statistical Methods in Medical Research* 2004; **13**: 525–38.

8 Berry DA, Cronin KA, Plevritis SK, *et al*. Effect of screening and adjuvant therapy on mortality from breast cancer. *New England Journal of Medicine* 2005; **353**: 1784–92.

9 Duffy SW, Chen HH, Tabar L, Day NE. 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* 1995; **14**: 1531–43.

10 Chen HH, Duffy SW. A Markov chain method to estimate the tumour progression rate from preclinical to clinical phase, sensitivity and positive predictive value for mammography in breast cancer screening. *The Statistician* 1996; **45**: 307–17.

11 Duffy SW, Chen HH, Tabar L, Fagerberg G, Paci E. Sojourn time, sensitivity and positive predictive value of mammography screening for breast cancer in women aged 40–49. *International Journal of Epidemiology* 1996; **25**: 1139–45.

12 Chen HH, Duffy SW, Tabar L, Day NE. Markov chain models for progression of breast cancer. Part I: tumour attributes and the preclinical screen-detectable phase.

*Journal of Epidemiology and Biostatistics* 1997; **2**: 9–23.
13 Chen HH, Duffy SW, Tabar L, Day NE. Markov chain models for progression of breast cancer. Part II: prediction of outcomes for different screening regimes. *Journal of Epidemiology and Biostatistics* 1997; **2**: 25–35.
14 Duffy SW, Day NE, Tabar L, Chen HH, Smith T. Markov models of breast tumor progression: some age-specific results. *Journal of the National Cancer Institute, Monographs* 1997; **22**: 93–7.
15 Straatman H, Peer PGM, Verbeek ALM. Estimating lead time and sensitivity in a screening program without estimating the incidence in the screened group. *Biometrics* 1997; **53**: 217–29.
16 Chen HH, Thurfjell E, Duffy SW, Tabar L. *et al.* Evaluation by Markov chain models of a non-randomised breast cancer screening programme in women aged under 50 years in Sweden. *Journal of Epidemiology and Community Health* 1998; **52**: 329–35.
17 Prevost TC, Rohan TE, Duffy SW, Chen HH, To T, Hill RD. Estimating sensitivity and sojourn time in screening for colorectal cancer: a comparison of statistical approaches. *American Journal of Epidemiology* 1998; **148**: 609–19.
18 Chen TH, Kuo HS, Yen MF, Lai MS, Tabar L, Duffy SW. Estimation of sojourn time in chronic disease screening without data on interval cases. *Biometrics* 2000; **56**: 167–72.
19 Chen TH, Yen MF, Lai MS, *et al.* Evaluation of a selective screening for colorectal carcinoma: the Taiwan Multicenter Cancer Screening (TAMCAS) project. *Cancer* 1999; **86**: 1116–28.
20 Myles JP, Nixon RM, Duffy SW, *et al.* Bayesian evaluation of breast cancer screening using data from two studies. *Statistics in Medicine* 2003; **22**: 1661–74.
21 Weedon-Fekjaer H, Vatten LJ, Aalen OO, Lindqvist B, Tretli S. Estimating mean sojourn time and screening test sensitivity in breast cancer mammography screening: new results. *Journal of Medical Screening* 2005; **12**: 172–8.
22 Weedon-Fekjaer H, Vatten LJ, Aalen OO, Lindqvist B, Tretli S. Estimating mean sojourn time and screening sensitivity using data on time since previous screening. *Journal of Medical Screening* 2008; **15**: 83–90.
23 Chien CR, Lai MS, Chen TH. Estimation of mean sojourn time for lung cancer by chest X-ray screening with a Bayesian approach. *Lung Cancer* 2008; **62**: 215–20.
24 Yen MF, Tabár L, Vitak B, Smith RA, Chen HH, Duffy SW. Quantifying the potential problem of overdiagnosis of ductal carcinoma *in situ* in breast cancer screening. *European Journal of Cancer* 2003; **39**: 1746–54.
25 Duffy SW, Agbaje O, Tabar L, *et al.* Overdiagnosis and overtreatment of breast cancer: estimates of overdiagnosis from two trials of mammographic screening for breast cancer. *Breast Cancer Research* 2005; **7**: 258–65.
26 Olsen AH, Agbaje OF, Myles JP, Lynge E, Duffy SW. Overdiagnosis, sojourn time, and sensitivity in the Copenhagen mammography screening program. *Breast Journal* 2006; **12**: 338–42.
27 Weedon-Fekjaer H, Lindqvist BH, Vatten LJ, Aalen OO, Tretli S. Breast cancer tumor growth estimated through mammography screening data. *Breast Cancer Research* 2008; **10**: R41.
28 Shen Y, Zelen M. Parametric estimation procedures for screening programmes: stable and nonstable disease models for multimodality case finding. *Biometrika* 1999; **86**: 503–15.
29 Shen Y, Zelen M. Robust modeling in screening studies: estimation of sensitivity and preclinical sojourn time distribution. *Biostatistics* 2005; **6**: 604–14.
30 Pinsky PF. Estimation and prediction for cancer screening models using deconvolution and smoothing. *Biometrics* 2001; **57**: 389–95.
31 Pinsky PF. An early- and late-stage convolution model for disease natural history. *Biometrics* 2004; **60**: 191–8.
32 Cong XJ, Shen Y, Miller AB. Estimation of age-specific sensitivity and sojourn time in breast cancer screening studies. *Statistics in Medicine* 2005; **24**: 3123–38.
33 Telesca D, Etzioni R, Gulati R. Estimating lead time and overdiagnosis associated with PSA screening from prostate cancer incidence trends. *Biometrics* 2008; **64**: 10–19.
34 van Oortmarssen GJ, Habbema JD, Lubbe JT, van der Maas PJ. A model-based analysis of the HIP project for breast cancer screening. *International Journal of Cancer* 1990; **46**: 207–13.

35 van Oortmarssen GJ, Habbema JD, van der Maas PJ, *et al.* A model for breast cancer screening. *Cancer* 1990; **66**: 1601–12.
36 Loeve F, Boer R, van Oortmarssen GJ, van Ballegooijen M, Habbema JD. The MISCAN-COLON simulation model for the evaluation of colorectal cancer screening. *Computers and Biomedical Research* 1999; **32**: 13–33.
37 van den Akker-van Marle ME, van Ballegooijen M, van Oortmarssen GJ, Boer R, Habbema JD. Cost-effectiveness of cervical cancer screening: comparison of screening policies. *Journal of the National Cancer Institute* 2002; **94**: 193–204.
38 Draisma G, Postma R, Schröder FH, van der Kwast TH, de Koning HJ. Gleason score, age and screening: modeling dedifferentiation in prostate cancer. *International Journal of Cancer* 2006; **119**: 2366–71.
39 T Tan SY, van Oortmarssen GJ, de Koning HJ, Boer R, Habbema JD. The MISCAN-Fadia continuous tumor growth model for breast cancer. *Journal of the National Cancer Institute Monographs* 2006; **36**: 56–65.
40 Arveux P, Wait S, Schaffer P. Building a model to determine the cost-effectiveness of breast cancer screening in France. *European Journal of Cancer Care* 2003; **12**: 143–53.
41 Lejeune C, Arveux P, Fagnani F, Bonithon-Kopp C, Faivre J. A simulation model for evaluating the medical and economic outcomes of screening strategies for colorectal cancer. *European Journal of Cancer Prevention* 2003; **12**: 77–84.
42 Rodríguez-Moranta F, Trapero-Bertran M, Castells A, *et al.* Endoscopic requirements of colorectal cancer screening programs in average-risk population. Estimation according to a Markov model. *Gastroenterologia Y Hepatologia* 2008; **31**: 405–12.
43 Jackson C, Sharples L, Thompson S, Duffy S, Couto E. Multistate Markov models for disease progression with classification error. *Statistician* 2003; **52**: 193–209.
44 Venable WM and Ripley BD. *Modern Applied Statistics with S-Plus* (3rd edn). Springer-Verlag, New York; 1999.
45 Paci E, Duffy SW. Modelling the analysis of breast cancer screening programmes: sensitivity, lead time and predictive value in the Florence District Programme (1975–1986). *International Journal of Epidemiology* 1991; **20**: 852–8.
46 Brekelmans CT, Westers P, Faber JA, Peeters PH, Collette HJ. Age specific sensitivity and sojourn time in a breast cancer screening programme (DOM) in The Netherlands: a comparison of different methods. *Journal of Epidemiology and Community Health* 1996; **50**: 68–71.
47 Foucher Y, Giral M, Soulillou JP, Daures JP. A flexible semi-Markov model for interval-censored data and goodness-of-fit testing. *Statistical Methods in Medical Research* 2008, doi: 10.1177/0962280208093889.
48 Foucher Y, Giral M, Soulillou JP, Daures JP. A semi-Markov model for multistate and interval-censored data with multiple terminal events. Application in renal transplantation. *Statistics in Medicine* 2007; **26**: 5381–93.
49 Foucher Y, Mathieu E, Saint-Pierre P, Durand JF, Daurès JP. A semi-Markov model based on generalized Weibull distribution with an illustration for HIV disease. *Biometrical Journal* 2005; **47**: 825–33.
50 Castelli C, Combescure C, Foucher Y, Daures JP. Cost-effectiveness analysis in colorectal cancer using a semi-Markov model. *Statistics in Medicine* 2007; **26**: 5557–71.

## Appendix: Derivation of the likelihood

We present the derivation of the log-likelihood for two screening rounds in the basic situation without opportunistic screening. We assume that all women have the same interval $t$ between their two screens and that follow-up for interval cancer after the second screen is equal to $t$. Details of calculations are available from the authors. Only one term is presented for interval cancers aggregated in a single class, to simplify the presentation. Terms for interval cancers are grouped by time since screens are

easily derived. The derivation is exact for the first and second screening rounds. The terms pertaining to the second round can be used as an approximation for subsequent rounds.

The derivation of the likelihood taking into account opportunistic screening is available from the authors (ZU or SWD). It is based on conditional probabilities given that a woman has or has not undergone opportunistic screening. It assumes opportunistic screening has the same sensitivity as organised screening. It uses a reasonable and helpful approximation: only the last two screenings are taken into consideration in conditional probabilities on the past history. Derivation is exact when sensitivity equals 1.

We denote $\{O_t\}$ to be the observed process at time $t$ and $\{E_t\}$ to be the real process. The process is observed at screening or at clinical diagnosis. We assume that a cancer discovered at screening is in a preclinical state. The observed state at screening might differ from the true state in case of a false-negative: a preclinical cancer can be wrongly considered as a non-disease state. The real process $\{E_t\}$ is assumed to be a homogeneous Markov process with instantaneous transition matrix $\Lambda$ and probability transition $P_\lambda(t) = \left(p_{i,j}(t)\right)_{i,j=1..3} = P[E_{a+t} = j \,|E_a = i]$ as defined in the material and methods section above.

We denote $a$, age at first screening, $t$ the interval between the screenings and $s$ the sensitivity of the mammography. Sensitivity links the observed state to the real state. For example, in the three-state model, $s$ is defined as: $P(O_t = 2 \,|E_t = 2) = s$ and $P(O_t = 1 \,|E_t = 2) = 1 - s$.

## Terms involved in the three-state model

- First screening (age $a$):

$$q_{f,1} = P(O_a = 2 \,|E_a \neq 3) = \frac{s.p_{12}(a)}{p_{11}(a) + p_{12}(a)}$$
$$q_{f,0} = P(O_a = 1 \,|E_a \neq 3) = 1 - q_{f,1}$$

- Interval cancers after first screening (interval $t$):

$$r_{f,1} = P(E_{a+t} = 3 \,|O_a = 1, E_a \neq 3) = p_{13}(t).\alpha_1(a) + p_{23}(t).\alpha_2(a)$$
$$r_{f,0} = P(E_{a+t} \neq 3 \,|O_a = 1, Ea \neq 3) = 1 - r_{f,1}$$

where:

$$\alpha_1(a) = P(E_a = 1 \,|O_a = 1, E_a \neq 3) = \frac{p_{11}(a)}{p_{11}(a) + (1-s).p_{12}(a)} \quad \text{and}$$
$$\alpha_2(a) = P(E_a = 2 \,|O_a = 1, E_a \neq 3) = 1 - \alpha_1(a)$$

- Second screen ($t$ after first screening)

$$q_{s,1} = P\left(O_{a+t} = 2 \,|O_a = 1, E_{a+t} \neq 3, E_a \neq 3\right) = \frac{s}{r_{f,0}}\left(p_{12}(t).\alpha_1(a) + p_{22}(t).\alpha_2(a)\right)$$
$$q_{s,0} = P\left(O_{a+t} = 1 \,|O_a = 1, E_{a+t} \neq 3, E_a \neq 3\right) = 1 - q_{s,1}$$

- Interval cancers after second screening (interval $t$)

$$\begin{aligned} r_{s,1} &= P\left(E_{a+2t} = 3 \,|O_{a+t} = 1, O_t = 1, E_{a+t} \neq 3, E_a \neq 3\right) \\ &= p_{13}(t).\beta_1(a,t) + p_{23}(t).\beta_2(a,t) \\ r_{s,0} &= P\left(E_{a+2t} \neq 3 \,|O_{a+t} = 1, O_t = 1, E_{a+t} \neq 3, E_a \neq 3\right) = 1 - r_{s,1} \end{aligned}$$

where:

$$\begin{aligned} \beta_1(a,t) &= P\left(E_{a+t} = 1 \,|O_{a+t} = 1, O_a = 1, E_{a+t} \neq 3, E_a \neq 3\right) \\ &= \frac{p_{11}(t).\alpha_1(a)}{p_{11}(t).\alpha_1(a) + (1-s)\,.\,(p_{12}(t).\alpha_1(a) + p_{22}(t).\alpha_2(a))} \\ \beta_2(a,t) &= P\left(E_{a+t} = 2 \,|O_{a+t} = 1, O_a = 1, E_{a+t} \neq 3, E_a \neq 3\right) = 1 - \beta_1(a,t) \end{aligned}$$

## Log-likelihood for the three-state model

We denote $n_1^f$, $n_0^f$ to be the number of women with and without cancer, respectively, detected at first screening and $k_1^f$, $k_0^f$ to be the number of women with and without interval cancers after the first screening and before the second screening. These numbers are denoted as $n_1^s$, $n_0^s$ and $k_1^s$, $k_0^s$, respectively, for the second screening. The log-likelihood is then expressed as:

$$\begin{aligned} \mathrm{Log}L(data\,|\lambda_{1,2}, \lambda_{2,3}, s) = {} & n_1^f * \log\left(q_{f,1}\right) + n_0^f * \log(q_{f,0}) + k_1^f * \log(r_{f,1}) \\ & + k_0^f * \log(r_{f,0}) + n_1^s * \log\left(q_{s,1}\right) + n_0^s * \log(q_{s,0}) \\ & + k_1^s * \log(r_{s,1}) + k_0^s * \log(r_{s,0}) + cst. \end{aligned}$$

The constant term is ignored and is not calculated since it does not affect maximisation.

## Terms involved in the five-state model

- First screening (age $a$):

$$q_{f,1} = P\left(O_a = 2 \,|E_a \neq \{4,5\}\right) = \frac{s.p_{12}(a)}{p_{11}(a) + p_{12}(a) + p_{13}(a)}$$
$$q_{f,2} = P\left(O_a = 3 \,|E_a \neq \{4,5\}\right) = \frac{s.p_{13}(a)}{p_{11}(a) + p_{12}(a) + p_{13}(a)}$$
$$q_{f,0} = P\left(O_a = 1 \,|E_a \neq \{4,5\}\right) = 1 - q_{f,1} - q_{f,2}$$

- Interval cancers after first screening (interval $t$):

$$r_{f,1} = P\left(E_{a+t} = 4 \,|O_a = 1, E_a \neq \{4,5\}\right) = p_{14}(t).\alpha_1(a) + p_{24}(t).\alpha_2(a)$$
$$r_{f,2} = P\left(E_{a+t} = 5 \,|O_a = 1, E_a \neq \{4,5\}\right) = p_{15}(t).\alpha_1(a) + p_{25}(t).\alpha_2(a) + p_{35}(t).\alpha_3(a)$$
$$r_{f,0} = P\left(E_{a+t} \neq \{4,5\} \,|O_a = 1, E_a \neq \{4,5\}\right) = 1 - r_{f,1} - r_{f,2}$$

where:

$$\alpha_1(a) = P\left(E_a = 1 \,|O_a = 1, E_a \neq \{4,5\}\right) = \frac{p_{11}(a)}{p_{11}(a) + (1-s)\,.\,(p_{12}(a) + p_{13}(a))},$$
$$\alpha_2(a) = P\left(E_a = 2 \,|O_a = 1, E_a \neq \{4,5\}\right) = \frac{(1-s)\,.p_{12}(a)}{p_{11}(a) + (1-s)\,.\,(p_{12}(a) + p_{13}(a))},$$
$$\alpha_3(a) = P\left(E_a = 3 \,|O_a = 1, E_a \neq \{4,5\}\right) = 1 - \alpha_1(a) - \alpha_2(a)$$

- Second screening ($t$ after first screening)

$$\begin{aligned} q_{s,1} &= P\left(O_{a+t} = 2 \,|O_a = 1, E_{a+t} \neq \{4,5\}, E_a \neq \{4,5\}\right) \\ &= \frac{s}{r_{f,0}}.\left(p_{12}(t).\alpha_1(a) + p_{22}(t).\alpha_2(a)\right) \\ q_{s,2} &= P\left(O_{a+t} = 3 \,|O_a = 1, E_{a+t} \neq \{4,5\}, E_a \neq \{4,5\}\right) \\ &= \frac{s}{r_{f,0}}.\left(p_{13}(t).\alpha_1(a) + p_{23}(t).\alpha_2(a) + p_{33}(t).\alpha_3(a)\right) \\ q_{s,0} &= P\left(O_{a+t} = 1 \,|O_a = 1, E_{a+t} \neq \{4,5\}, E_a \neq \{4,5\}\right) = 1 - q_{s,1} - q_{s,2} \end{aligned}$$

- Interval cancers after second screening (interval $t$)

$$\begin{aligned} r_{s,1} &= P\left(E_{a+2t} = 4 \,|O_{a+t} = 1, O_a = 1, E_{a+t} \neq \{4,5\}, E_a \neq \{4,5\}\right) \\ &= p_{14}(t).\beta_1(a,t) + p_{24}(t).\beta_2(a,t) \\ r_{s,2} &= P\left(E_{a+2t} = 5 \,|O_{a+t} = 1, O_a = 1, E_{a+t} \neq \{4,5\}, E_a \neq \{4,5\}\right) \\ &= p_{15}(t).\beta_1(a,t) + p_{25}(t).\beta_2(a,t) + p_{35}(t).\beta_3(a,t) \\ r_{s,0} &= P\left(E_{a+2t} \neq \{4,5\} \,|O_{a+t} = 1, O_a = 1, E_{a+t} \neq \{4,5\}, E_a \neq \{4,5\}\right) \\ &= 1 - r_{s,1} - r_{s,2} \end{aligned}$$

where:

$$\beta_1(a,t) = P\left(E_{a+t} = 1 \,|O_{a+t} = 1, O_a = 1, E_{a+t} \neq \{4,5\}, E_a \neq \{4,5\}\right),$$
$$\beta_2(a,t) = P\left(E_{a+t} = 2 \,|O_{a+t} = 1, O_a = 1, E_{a+t} \neq \{4,5\}, E_a \neq \{4,5\}\right),$$
$$\beta_3(a,t) = P\left(E_{a+t} = 3 \,|O_{a+t} = 1, O_a = 1, E_{a+t} \neq \{4,5\}, E_a \neq \{4,5\}\right).$$

And:

$$\beta_1(a,t) = \frac{p_{11}(t).\alpha_1(a)}{\begin{array}{l}[p_{11}(t) + (1-s)\,(p_{12}(t) + p_{13}(t))]\,.\alpha_1(a) + (1-s)\,(.p_{22}(t) \\ + p_{23}(t))\,.\alpha_2(a) + (1-s)\,p_{33}(t).\alpha_3(a)\end{array}},$$

$$\beta_2(a,t) = \frac{(1-s)\,.\,(p_{12}(t).\alpha_1(t) + p_{22}(t).\alpha_2(t))}{\begin{array}{l}[p_{11}(t) + (1-s)\,(p_{12}(t) + p_{13}(t))]\,.\alpha_1(a) + (1-s)\,(.p_{22}(t) \\ + p_{23}(t))\,.\alpha_2(a) + (1-s)\,p_{33}(t).\alpha_3(a)\end{array}},$$

$$\beta_3(a,t) = 1 - \beta_1(a,t) - \beta_2(a,t)$$

## Log-likelihood for the five-state model

We denote $n_0^f$, $n_1^f$, $n_2^f$ to be the number of women without cancer, with cancer node negative and with cancer node positive, respectively, detected at first screening. We denote $k_0^f$, $k_1^f$, $k_1^f$ to be the number of women without interval cancer, with interval cancer node negative and with interval cancer node positive, after the first screening and before the second screening. These numbers are denoted as $n_0^s$, $n_1^s$, $n_2^s$ and $k_0^s$, $k_1^s$, $k_2^s$, respectively, for the second screening. The log-likelihood is then expressed as:

$$\begin{aligned}\mathrm{Log}L(data|\lambda_{1,2},\lambda_{2,3},\lambda_{2,4},\lambda_{3,5},s) = {} & n_1^f * \log(q_{f,1}) + n_2^f * \log(q_{f,2}) + n_0^f * \log(q_{f,0}) \\ & + k_1^f * \log(r_{f,1}) + k_2^f * \log(r_{f,2}) + k_0^f * \log(r_{f,0}) \\ & + n_1^s * \log(q_{s,1}) + n_2^s * \log(q_{s,2}) + n_0^s * \log(q_{s,0}) \\ & + k_1^s * \log(r_{s,1}) \; + k_2^s * \log(r_{s,2}) + k_0^s * \log(r_{s,0}) \\ & + cst\end{aligned}$$