Cancer Epidemiology 35 (2011) 235–242

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Modelling the effect of breast cancer screening on related mortality using French data

Z. Uhry [a,*], G. Hédelin [b], M. Colonna [c], B. Asselain [d], P. Arveux [e], C. Exbrayat [f], C. Guldenfelds [g], P. Soler-Michel [h], F. Molinié [i], B. Trétarre [j], A. Rogel [a], I. Courtial [h], A. Danzon [k], A.V. Guizard [l], R. Ancelle-Park [m], D. Eilstein [n], S. Duffy [o]

[a] *Département des maladies chroniques et des traumatismes, Institut de veille sanitaire, 12, rue du Val d'Osne, 94 415 St-Maurice Cedex, France*
[b] *Registre des cancers du Bas-Rhin, Laboratoire d'épidémiologie et de santé publique, EA3430, Université de Strasbourg, Strasbourg, France*
[c] *Registre des cancers de l'Isère, Meylan, France*
[d] *Service de biostatistiques, Institut Curie, Paris, France*
[e] *Registre des cancers de Côte d'Or, Centre Leclerc, EA 4184, Dijon, France*
[f] *Structure de gestion du dépistage de l'Isère, Meylan, France*
[g] *Structure de gestion du dépistage du Bas-Rhin, Strasbourg, France*
[h] *Structure de gestion du dépistage du Rhône, Lyon, France*
[i] *Registre des cancers de Loire-Atlantique, Nantes, France*
[j] *Registre des cancers de l'Hérault, Montpellier, France*
[k] *Registre des cancers du Doubs, Besançon, France*
[l] *Registre des cancers du Calvados, Caen, France*
[m] *Direction générale de la santé, Ministère de la santé, Paris, France*
[n] *Département santé environnement, Institut de veille sanitaire, St-Maurice, France*
[o] *Department of epidemiology, mathematics and statistics, Cancer Research UK, London, United Kingdom*

ARTICLE INFO

Article history:
Accepted 31 October 2010
Available online 14 December 2010

Keywords:
Breast cancer screening
Mortality
Markov model
Relative survival
Natural history

ABSTRACT

Introduction: This study aimed at modelling the effect of organized breast cancer screening on mortality in France. It combined results from a Markov model for breast cancer progression, to predict number of cases by node status, and from relative survival analyses, to predict deaths. The method estimated the relative risk of mortality at 8 years, in women aged 50–69, between a population screened every two years and a reference population. *Methods*: Analyses concerned cases diagnosed between 1990 and 1996, with a follow-up up to 2004 for the vital status. Markov models analysed data from 3 screening programs (300,000 mammographies) and took into account opportunistic screening among participants to avoid bias in parameter's estimates. We used survival data from cancers in the general population (n = 918, 7 cancer registries) and from screened cancers (n = 565, 3 cancer registries), after excluding a subgroup of screened cases with a particularly high survival. Sensitivity analyses were performed. *Results*: Markov model main analysis lacked of fit in two out of three districts. Fit was improved in stratified analyses by age or district, though some lack of fit persisted in two districts. Assuming 10% or 20% overdiagnosed screened cancers, mortality reduction was estimated as 23% (95% CI: 4, 38%) and 19% (CI: −3, 35%) respectively. Results were highly sensitive to the exclusion in the screened cancers survival analysis. Conversely, RR estimates varied moderately according to the Markov model parameters used (stratified by age or district). *Conclusion*: The study aimed at estimating the effect of screening in a screened population compared to an unscreened control group. Such a control group does not exist in France, and we used a general population contaminated by opportunistic screening to provide a conservative estimate. Conservative choices were systematically adopted to avoid favourable estimates. A selection bias might however affect the estimates, though it should be moderate because extreme social classes are under-represented among participants. This modelling provided broad estimates for the effect of organized biennial screening in France in the early nineteen-nineties. Results will be strengthened with longer follow-up.

1. Introduction

The main objective of breast cancer screening is to reduce breast cancer mortality [1]. Overviews of randomized trials suggested regular screening would reduce breast cancer mortality

* Corresponding author. Tel.: +33 1 55 12 54 22; fax: +33 1 41 79 68 11.
E-mail address: z.uhry@invs.sante.fr (Z. Uhry).

1877-7821/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.canep.2010.10.009

by 25–30% in screened women [1–3]. Some authors, however, contested conclusive evidence from these trials [4]. Evaluating the effect of breast cancer service screening on mortality is a crucial issue for health authorities. Evaluation is difficult though outside the experimental context and a variety of methods have been used [5–18]. In France, pilot programs were first initiated in a few districts (*départements*) in 1990; a national program was later initiated and finally extended to all districts in 2004. Opportunistic screening is common and has developed in parallel to service screening programs. This study aimed at evaluating the effect of breast cancer screening on related mortality in France, comparing a screened population to an unscreened control group. We revised a method initially proposed by Chen and Duffy [13,14].

2. Material and methods

The method combines results from a Markov model for the progression of cancer and from relative survival analyses, using node status as prognostic factor. It requires survival from cancers diagnosed at screening and from a reference population. The Markov model is used to predict the numbers of cases and relative survival to predict deaths. The method estimates the relative risk (RR) of mortality between a population screened every two years and a reference population.

There is no proper control group in France to use as reference population, due to the concomitant development of opportunistic screening. We thus considered two alternatives: either cancers from the general population, but some were screened cancers outside the program, or clinical cancers diagnosed due to symptoms, but they represent a specific population (essentially women not undergoing opportunistic screening), that we will name "unscreened reference population". These two references, however, provide bounds for the RR we want to estimate, comparing a screened population to an unscreened control group.

2.1. Data and population

Analyses concerned invasive breast cancers in women aged 50–69. The study period was restricted to 1990–96 to ensure 8-years follow-up in the survival analyses. The data sources are presented in Tables 1 and 2. The Markov model analysed data from screening programs (detection rates at first and subsequent screens and interval cancer rates). Three districts involved in the pilot programs, covered by a cancer registry or with exhaustive monitoring of interval cancers, could be analyzed (Bas-Rhin, Isère, Rhône). Data pertained to 180,000 women and 300,000 mammography episodes (one-view, double-reading). Screening targeted women aged 50–69, except in Bas-Rhin (aged 50–64), and women were invited every 2, 2.5 and 3 years respectively in Bas-Rhin, Isère and Rhône.

For each reference population (general population or unscreened population), only one source provided adequate data: a sample of cases diagnosed in 1990 issued from the French cancer registries network (Francim-breast) and cases from the cancer registry of Loire-Atlantique, which monitored detection mode and identified cancers detected on clinical signs (clinical cancers).

2.2. Multi-state Markov model for the progression of cancer

We used a 5-state homogeneous Markov model to describe the natural progression of breast cancer, from no disease to preclinical asymptomatic phase and finally to clinical phase, according to node status. Women start without disease (state 1, no disease). Cancers not detectable by screening are assimilated to the 'no disease state'. Women enter in state 2 (preclinical cancer, node negative) when they get a cancer detectable by screening. From state 2, the cancer can either spread to lymph node staying asymptomatic (sate 3, preclinical, node positive) or become symptomatic (state 4, clinical, node negative). From state 3, the node positive cancer will finally become symptomatic (sate 5, clinical, node positive). The parameters of this model are the transition rates $\lambda_{i,j}$ from state i to state j. $\lambda_{1,2}$ represents the incidence rate of the preclinical disease. Markov model implies that sojourn times in the two preclinical states (node negative and node positive) both follow an exponential distribution and that they are independent. Let's first consider a screening program in the absence of opportunistic screening. Tumours may be detected by screening or diagnosed symptomatically between rounds of screening (interval cancer). Observed data consists of detection rates at first and subsequent screens and interval cancer rates, by node status. Screened cancers are assumed in preclinical states while interval cancers are assumed in clinical states. Screening rhythm must be specified for each district. Observations depend both on the transition rates (natural history of the cancer) and on the sensitivity of the mammography, which are estimated jointly by maximum likelihood. Cancers with unknown node status are integrated in the likelihood assuming that unknown node status is independent of the true node status.

Table 1
Data analyzed in Markov models (sources screening management centers). Node status according to detection mode.

District	Cancers at first screen			Cancers at subsequent screens			Interval cancers		
	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

[a] Number of cancers with known node status.
[b] Proportion of node negative cancers among cancers with known node status.
[c] Proportion of cancers with unknown node status.

Table 2
Data analyzed in relative survival analyses[a,b] (sources cancer registries). Population, age and node status distribution.

Population	Sources	Period	Age class (%)		Node status	Number of cases
			50–59	60–69	%pN−	
Screened cancers	Bas-Rhin, Isère, Hérault	1990–96	42	58	72	731
General population	7 registries[c]	1990	46	54	58	918
Clinical cancers[d]	Loire-Atlantique	1991–95	47	53	52	811

[a] Node status (pN) was replaced by the clinical node status (N) for women without curage. In addition, pN status for cancers node negative after adjuvant chemotherapy in Loire-Atlantique were also replaced by the clinical node status N.
[b] End points between 2002 and 2004 according to data; analyses censored to 8 years.
[c] Bas-Rhin, Isère, Hérault, Calvados, Doubs, Somme, Tarn.
[d] Cancers diagnosed due to clinical signs or symptoms ("unscreened population").

In the data analyzed however, additional opportunistic screening (OS) was common among participants, especially before first organised screening. Observed data thus also depend on opportunistic screening practices. For each district, we made assumptions on: (1) proportion of women undergoing OS before 1st organised screen and (2) proportion of women undergoing OS after participating to organised screen and before next invitation. These assumptions were derived from self-reported data on previous screens and from studies on detection mode of interval cancers. The likelihood derivation integrated these assumptions by using conditional probabilities on whether a woman has or has not undergone opportunistic screening. We assumed opportunistic screening had the same sensitivity as organised screening.

Districts were analysed jointly, assuming different incidence and common parameters otherwise (same progression of cancer after its inception and same sensitivity of the mammography). Analyses were realised using S-Plus and likelihood derivation was detailed in a specific paper [19]. In order to perform next a sensitivity analysis of the RR estimate, age-specific (50–59 and 60–69) and district-specific analyses were also performed, as well as a sensitivity analysis regarding the assumptions adopted on opportunistic screening.

2.3. Relative survival analyses

Relative survival considers the excess in all cause mortality, in women with breast cancer, compared to expected mortality in the general population [20]. Relative survival models can be performed using the likelihood maximization procedure used in generalized linear models [21]. We applied this approach using SAS. Relative survival analyses were performed separately for screened cancers and reference populations. The final model only includes node status, but age (5-years class) and district effects were first tested. Analyses were censored at 8 years to ensure identical follow-up in all districts. The excess mortality rate according to time since diagnosis was assumed constant by interval (4 classes). These intervals were tested against 1-year intervals and were satisfactory. Log-linearity assumption of the model was also verified.

We used female mortality tables by district, annual age and year to calculate the expected general mortality among cases. An implicit assumption when applying all cause mortality from the general population is that cases are representative of the general population in terms of mortality from other causes than the disease in question. For screened and clinical cancers, this assumption may not be verified, giving rise to potential bias in survival estimates and, as a consequence, to overestimation of the benefit from screening. We thus applied correction factors to expected mortality rates in relative survival analyses of these two populations (0.8 and 1.2 respectively). These factors were derived from the highest and lowest comparative mortality rate ratio observed across social classes [22]. Although these social inequalities dot not estimate correction factors specifically for these two populations, they seemed reasonably conservative.

2.4. Estimation of the relative risk of mortality between a screened and a reference population

The parameters estimated in the Markov model analysis are then used to predict yearly the number of cases according to node status and detection mode in a population screened every two years (Table 4). We assumed same preclinical incidence in the screened and the reference populations (2.4‰, central value among district estimates). For the reference population, we used the directly observed proportion of node negative tumours. The relative risk (RR) was then estimated after 8 years for a population screened every two years compared to a reference population. Survival from screened cancer is used to predict deaths in screened cancers while survival from the reference population was used to predict deaths in interval cancers and in the reference population. Calculation of the RR is detailed in Appendix A. Deaths were predicted to a fixed date, at the end of the 8-year period, to avoid lead time bias. This means that the term used to predict death depends on the date of diagnosis. The earlier the cancer is detected, the longer is the term used to predict deaths. Table 4 presents the number of predicted cases from the Markov model, year by year, and term used to predict deaths.

The RR estimate depends on parameters from three independent analyses (Markov model, survival of screened cancers and of the reference population). We used the empirical distribution obtained from 4000 simulations to estimate the mean RR and its 95% confidence interval. Simulations are based on the covariance matrices of the parameters, assuming each follows a multivariate normal distribution. Calculations were carried out using S-Plus.

2.5. Correcting for overdiagnosis

Overdiagnosis refers to the detection of cancer by screening which would never have become symptomatic in the woman's lifetime otherwise. Its magnitude is not known precisely [23]. Overdiagnosis may affect all results. A sensitivity analysis of the RR estimate is presented, correcting respectively for 10% or 20% screened cancers overdiagnosed. It is based on a simple correction of the node negative screened cancers survival. This correction was first validated on a simulation study (not presented).

3. Results

3.1. Progression model

Table 1 presents the data analysed for the progression model. Note that the proportion of node negative cancers was higher at first screen compared to subsequent screens in Bas-Rhin and Isère,

Table 3
Observed[a] and predicted number of cases, 50–69 years (multi-district analysis).

Detection mode	Bas-Rhin			Isère			Rhône		
	Obs	Pred	Chi[b]	Obs	Pred	Chi[b]	Obs	Pred	Chi[b]
First screen. pN−	160	158	0.0	156	139	2.0	351	334	0.9
First screen. pN+	56	68	2.1	50	55	0.5	140	132	0.5
Subsequent screen. pN−	106	130	4.4	34	47	3.5	133	146	1.2
Subsequent screen. pN+	48	36	3.8	20	14	3.1	52	45	1.1
Interval cancer. pN−	93	85	0.5	93	92	0.0	327	346	1.0
Interval cancer. pN+	64	49	4.3	37	49	3.1	187	189	0.0
Total	527	527	15.1[c]	390	396	12.2[c]	1 190	1 192	4.7[c]

[a] Fit is presented restricted to cancer with known nodes status.
[b] $Chi = (Obs - Pred)^2/Pred$.
[c] Sum of Chi across detection mode.

Table 4
Predicted number of breast cases and term used for predicting related deaths by year. Illustration for 100,000 women assuming a preclinical incidence of 2.4‰.

Year	Screened population						Reference population[a]		
	Screened cancers			Interval cancers			Clinical cancers		
	pN−[b]	pN+[b]	Term[c]	pN−[b]	pN+[b]	Term[c]	pN−[b]	pN+[b]	Term[c]
1	389	186	8	36	23	7.5	125	115	7.5
2	–	–	–	66	42	6.5	125	115	6.5
3	271	76	6	33	14	5.5	125	115	5.5
4	–	–	–	63	35	4.5	125	115	4.5
5	269	75	4	32	14	3.5	125	115	3.5
6	–	–	–	63	35	2.5	125	115	2.5
7	268	75	2	32	14	1.5	125	115	1.5
8	–	–	–	63	35	0.5	125	115	0.5
Total[d]	1.609	–	600	–	1.920	–			

[a] Assuming 52% cancers without node involvement (reference unscreened population, i.e. clinical cancers from Loire-Atlantique).
[b] Number of predicted cases according to node status (pN−: node negative. pN+: node positive).
[c] Term (in years) used to predict related deaths among cases. The term is at mid-year for clinical cancers (interval cancer or reference population) since they occur all year while we assumed cancers are screened at the beginning of the year.
[d] Total number of cases. It is higher in the screened population ($n = 1.609 + 600 = 2.209$) compared to the reference population (1.920) because some screened cancers would not have been diagnosed yet at the end of the 8 years in the absence of screening.

which is unexpected in screening results. Average duration of the preclinical phase was estimated as 2.8 years (CI: 2.3, 3.5) and sensitivity of the mammography as 86% (79, 92%) [19]. Table 3 presents a summary table of observed and predicted number of cases, according to detection mode, node status and district, for all age analyses (50–69). The overall fit is poor for all age analysis (sum of chi-square measures $Chi^2 = 32.0$, df = 11, threshold = 19.7). The lack of fit mainly concerns the cancers detected at subsequent screens in Bas-Rhin and Isère, with over-estimation of proportion of node negative cancers. It also concerned node positive interval cancers, which were underestimated in Bas-Rhin and over-estimated in Isère. In age-specific analyses (data not shown), overall fit was better (22.9 in 50–59 and 21.1 in 60–69), though lack of fit persisted for proportion of node negative cancers at subsequent screens in Bas-Rhin and Isère. In district-specific analyses, this lack of fit also persisted while interval cancers were correctly predicted.

Table 4 presents the prediction over 8 years obtained from the model for a population screened every two years. The model predicted 72% node negative tumours in a screened population, compared with 52% observed in an unscreened population (clinical cancers from Loire-Atlantique) and 58% in the general population (Francim-breast).

3.2. Survival analyses

Table 2 presents the data analyzed. Age (5-years classes) and district effects, adjusted for node status, were tested on the original data before correction of the expected mortality. They were not significant in the general population (Francim-breast) or in clinical cancers (all p greater than 0.4). In screened cancers, we excluded cases aged over 60 in Bas-Rhin ($n = 166$) because these women had a particularly high survival, in order to provide conservative estimates. After this exclusion, 565 cases were included in the analysis and 46% were aged 60–69. The district effect was not significant ($p = 0.6$) neither age effect ($p = 0.2$) after this exclusion. Fig. 1 presents the relative survival estimates finally obtained, after correcting the expected mortality for screened and clinical cancers.

3.3. Estimation of the relative risk of mortality

Table 5 presents the RR estimates for both reference populations and corrected for overdiagnosis. As expected, mortality reduction was lower using the general population as reference rather than unscreened population. It was estimated as 23% (4, 38%) and 19% (−3, 35%) assuming 10% or 20% overdiagnosed screened cancers (reference general population).

The RR estimate was highly sensitive to the screened cancers survival. Without exclusion of cases aged over 60 in Bas-Rhin, the mortality reduction would be estimated as 32% (15, 44%) and 28% (9, 41%) assuming 10% or 20% overdiagnosed screened cancers (reference general population). We tested two alternatives to represent a general population: restriction of the Francim-breast sample to the same districts as included in the screened cancer analysis, or all cancers from Loire-Atlantique. The RR were similar using these two alternatives (0.74 and 0.75 respectively, uncorrected) to the estimate obtained using the Francim-breast sample (0.73).

Table 6 presents the RR obtained when varying the assumptions on opportunistic screening (OS) in the Markov model analyses (reference general population, uncorrected). The percentage of women undergoing OS was increased or decreased simultaneously in the three districts. RR estimates were only slightly modified. Table 7 presents RR's based on Markov model analyses stratified by district or age. RR's exhibited moderate variations. Varying solely the sensitivity of the mammography, from 75% to 95%, would lead to RR's ranging from 0.78 to 0.70.

We explored RR's obtained without correcting the expected mortality in the relative survival analyses of screened and clinical cancers. We found estimates of 0.70 and 0.61 respectively using general population (Francim-breast) or unscreened population (clinical cancers, Loire-Atlantique) as a reference. Correcting expected mortality may avoid a slight bias in favour of screening.

4. Discussion

This paper presents a modelling exercise to estimate the relative risk of breast cancer mortality after 8 years between a population screened with mammography every two years and a reference population. It combines results from a Markov model for breast cancer progression, to predict number of cases by node status, with those from relative survival analyses, to predict deaths. Relative survival represents an interesting alternative to cause-specific mortality in cancer screening evaluation. Screened cancers survival may be used, in this method, without bias due to lead time. Mortality reduction was estimated as 23% (4, 38%) and 19% (−3, 35%) assuming 10 or 20% overdiagnosed screened cancers. Our modelling is a simplified description of a complex reality and necessarily incorporates assumptions, but it enables

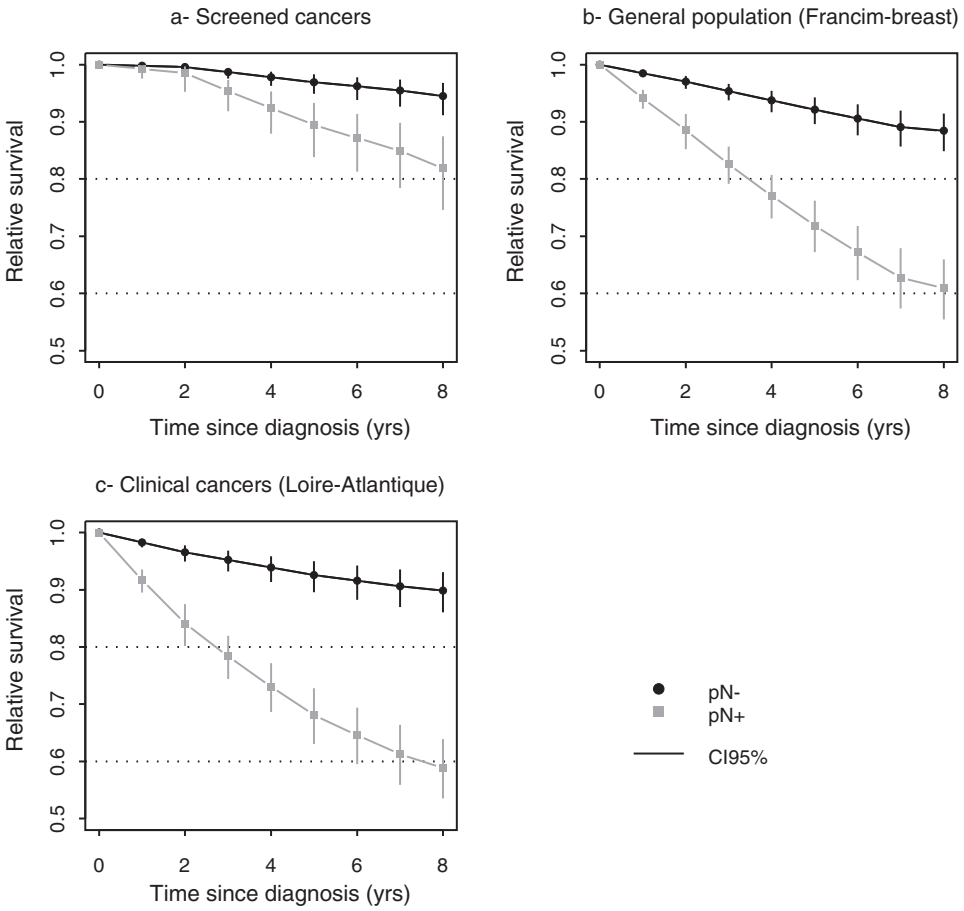

Fig. 1. Breast cancer relative survival cancer according to node status (pN−: negative, pN+: positive). (a) Screened cancers[a] (Bas-Rhin, Isère, Hérault), (b) General population (Francim-breast), (c) Unscreened population[a] (Clinical cancers from Loire-Atlantique).
[a]Correction factors of 0.8 and 1.2 were applied to expected mortality rates for screened cancers and clinical cancers respectively.

Table 5
Estimated relative risk of mortality (RR) and confidence intervals (CI), corrected for overdiagnosis.

Correction for overdiagnosis	Reference	
	General population[a]	Unscreened population[b]
None	0.73 (0.60, 0.91)	0.64 (0.53, 0.80)
10%[c]	0.77 (0.62, 0.96)	0.67 (0.55, 0.84)
20%[c]	0.81 (0.65, 1.03)	0.71 (0.60, 0.90)

[a] Sample from 7 cancer registries (Francim-breast) contaminated by opportunistic screening.
[b] Clinical cancers from Loire-Atlantique.
[c] Assumed proportion overdiagnosed screened cancers in the survival analyses.

Table 6
Estimated relative risk of mortality (RR) depending on the assumptions adopted in the Markov model about opportunistic screening (OS).

Assumption about OS	RR
Main analysis[a]	0.73
Proportion of women undergoing OS before 1st participation decreased by 10%[b]	0.72
Proportion of women undergoing OS before 1st participation increased by 10%[b]	0.76
Proportion of women undergoing OS between two screens decreased by 10%[b]	0.72
Proportion of women undergoing OS between two screens increased by 10%[b]	0.74

[a] Assumption adopted in main analysis.
[b] Proportions are increased or decreased in all districts simultaneously.

summarizing information from all available data, taking into account opportunistic screening. Assumptions, however, are explicit and can be discussed and sensitivity analyses allow identifying crucial issues in the relative risk estimation.

Estimates of the mean sojourn time in preclinical phase (MST) and of the sensitivity of the mammography from Markov model were consistent with international literature [17,24–27]. Proportions of node negative tumours at subsequent screens were overestimated in two districts. This lack of fit is due to an odd configuration in our data, in that the observed proportion of node negative tumours was smaller at subsequent screens compared to the first screen in these two districts. Actually, the model is unable to predict such a configuration and provides predictions that conform more to expected screening results. For this reason, we

Table 7
Estimated relative risk of mortality (RR) depending on the Markov model parameters used using same survival estimates[a].

Markov model	RR
Main analysis[b] (50–69 years)	0.73
50–59 years[b]	0.76
60–69 years[b]	0.72
Bas-Rhin[c]	0.72
Isère[c]	0.69
Rhône[c]	0.77

[a] Survival estimates used are the same in all analyses.
[b] Multi-districts analyses (Bas-Rhin, Isèren Rhône).
[c] All ages analyses (50–69 years; except Bas-Rhin 50–64 years).

think this lack of fit was inevitable and does not invalidate our results. Note that in the successive evaluations of the French national screening program, proportions of node negative tumours were always higher at subsequent screens compared to first screen. A lack of fit was also observed for node positive interval cancers (over-estimated in Isère, under-estimated in Bas-Rhin), but it was attenuated in age-specific analyses and disappeared in district-specific analyses, suggesting it is due to the overall multi-district analysis. Sensitivity analyses showed this point had a moderate impact on the RR estimate. Our analyses also revealed a more serious limitation of the 5-state Markov model, which is detailed in a dedicated paper [19]. Briefly, we initially planned to use also the model to predict the proportion of node negative tumours in an unscreened population. But this proportion was a little underestimated using the model, even when integrating data from Loire-Atlantique. We finally used directly the proportion observed in the reference population. This alternative led to the best possible fit on screening data and avoided a bias in favour of screening. A better solution would be to adopt more flexible models. This implies research developments and several orientations may be considered, *e.g.* semi-Markov models [28] or tumour growth models [29].

As expected, the mortality reduction estimate was higher using an unscreened population (clinical cancers only) than general population reference, which is contaminated by opportunistic screening. These two estimates provide bounds for the RR estimate we would obtain using an unscreened control group. Clinical cancers reference is anti-conservative since it represents women not undergoing opportunistic screening, who probably tend to be diagnosed later than other women even if the latter were not screened. Conversely, general population reference is conservative because it is contaminated by opportunistic screening and should be preferred. We consider the RR using general population reference as a conservative estimate for the effect of organised screening in a screened population compared to an unscreened control group.

The RR estimate was highly sensitive to the survival of the screened cancers. And actually, its wide confidence intervals are mainly due to the large variance of this survival (data not shown). Conversely, RR estimates varied only moderately according to Markov model parameters used (age or district stratified analyses, or when varying assumptions about opportunistic screening). RR estimates also varied moderately when assumptions ranged from 10 to 20% of overdiagnosed screened cancers. They were little modified when we omitted correction of the expected mortality in the relative survival analyses. Finally, they were similar when using alternative general populations from the same period.

Selection bias may affect the RR estimate in favour of screening, since it compares participants to a reference population. The method proposed by Duffy et al. to correct for selection bias is not adapted to the predicted mortality indicator we used and could not be applied [30]. However, selection bias does not concern all terms involved in the RR estimation. Survival of the reference population is also used in the screened population to predict deaths from interval cancers, which represent 40% of the deaths in this screened population. In addition, extreme social classes are under-represented among participants in France, since women from affluent social classes are often screened outside the program while women from deprived social classes are often unscreened [31]. Selection bias should therefore remain moderate using general population reference, and it may be partly compensated by contamination. We cannot rule out, however, the possibility that it explains some of the mortality reduction estimated. In order to avoid selection bias, one might think using a single survival (reference population) for all cancers, whatever the detection mode. This means translating the gain in the proportion of node negative tumours due to screening, in terms of mortality. This solution may seem appealing but actually can only provide a positive result, since screening necessarily increases the proportion of node negative tumours, although it does not necessarily reduces mortality.

The method allows analysing data from several districts jointly at each step (progression model, survival of screen-detected cancers, survival of cancers in the reference population). RR is then estimated by combining results from these multi-districts analyses. Theoretically, we would prefer to estimate the relative risk in each district, test for heterogeneity, and finally estimate the overall RR. However, this would restrict the analysis to Isere and Bas-Rhin, and estimates would actually be meaningless because of their high imprecision. Analysing jointly data at each step allows both using all available information and consolidating the RR estimate. In addition, district-variability can be explored. In the Markov models, we reasonably assumed the progression of cancer was identical in all districts. We also assumed sensitivity of the mammography was the same, which is more questionable. Only moderate variations were observed, however, in the RR estimates in the sensitivity analyses. In the survival analysis of cancers from the general population, district effect was not observed ($p = 0.9$). In the survival analysis of the screened cancers, it faded after excluding elderly cases of Bas-Rhin ($p = 0.6$). These cases, which had a particularly high survival, were excluded in order to provide conservative estimate, although this probably lead to pessimistic lower bounds for the mortality reduction estimates. Longer follow-up and additional data in the survival analysis of screened cancers will help strengthen the results and shrink confidence intervals.

The magnitude of overdiagnosis due to breast cancer screening is an open question [22] and estimations vary widely according to studies and methods [32]. Breast cancer incidence trends raised concern about overdiagnosis [33], though they are difficult to interpret [34]. In particular, the role of hormone replacement therapy (HRT) is in question [35–37]. We tested assumptions from 10 to 20% overdiagnosed screened cancers in the sensitivity analyses. Ten percent represents a moderate level, consistent with some of the lower estimations [38,39]. Twenty percent represents a high level of overdiagnosis, though even higher estimations have been published [40–42]. However, plausibility of higher level may be discussed regarding our data. Forty percent of screened cancer overdiagnosed is undoubtedly an upper bound. Because overdiagnosis arguably only applies to node negative tumours, this assumption would imply that 55% of progressive screened cancers are node negative. This is incompatible with the 52% observed among clinical cancers. Assuming 30% overdiagnosed would imply 60% node negative progressive screened cancers. This is still unlikely, because the preclinical node positive phase would have to be longer, on average, than the preclinical node negative phase. We would not expect node positive tumours to grow especially slower after they had spread to lymph nodes. RR estimates corrected for 10% and 20% overdiagnosis would thus represent a reasonable range of plausible values.

Another aspect of overdiagnosis relates to the progression of *in situ* towards invasive carcinoma, which is not systematic. The model does not account for these different tumour behaviours. Only a minor difference was obtained when including *in situ* carcinoma in the analysis (RR = 0.69, general population reference, uncorrected). This result is slightly biased in favour of screening because some *in situ* detected by screening would not progress and are overdiagnosed. Results of both analyses are consistent actually, and the choice of our main analysis excluding *in situ* is reinforced.

The relative risk we estimated represents a plausible magnitude the effect of organized biennial screening in the early nineteen-

nineties. The actual effect probably varies around this value according to district, age, and screening rhythm. Screening modalities, as well as breast cancer treatments, have evolved since and our estimates do not apply to the current national breast cancer screening program. In particular, current program uses two-view mammography, instead of one-view in the pilot programs, which has decreased the interval cancer rates [43]. This may have improved the long-term effect on mortality.

Various approaches have been used to evaluate the effect of breast cancer services screening on mortality [44]: case–control studies [l5,16], observational cohort studies [5–8] and modelling approaches [9], whether likelihood-based [13,14] or simulation-based [10–12,18,45]. Simulation-based models are flexible and useful to realise cost-effectiveness analyses, but they are deterministic and do not account for the random variability of the results. Our modelling is less flexible, but parameters are estimated by maximum likelihood and confidence intervals are easily derived. They properly account for the random variability of the different analyses (Markov model and survival analyses). In the French context, observational studies did not seem appropriate due to the high contamination with opportunistic screening. Case–control studies require retracing screening history from deaths certificates, which is not feasible. Conversely, modelling seemed relatively adapted and was essential to summarize information from various pilot programs. The modelling adopted, however, impacts the results. A modelling consortium presented results from seven models of the effect of screening in the United-States on national annual breast cancer mortality rates (30–79 years) [9,11]. The estimated reduction in mortality rates in 2000 due to screening ranged from 7 to 23% across the seven models, with a median of 15%, illustrating the results also depend on the method used.

Our results are consistent with a former French cost-effectiveness study in Bas-Rhin (12% cumulative mortality reduction with 54% participation) [12]. In international studies, mortality reduction in participants to service screening programs, corrected for selection bias, ranged from 28% to 48% [5–8,15,16,44]. Our central estimates are lower, though similar to a model estimate from Catalonia [18]. French pilot services screening may have been less effective, but systematic conservative choices may also have contributed to these lower estimates.

5. Conclusion

This study presents a revised method to model the effect on mortality associated with breast cancer screening using French data. The method combines results from a Markov progression model and relative survival analyses. Modelling enabled us to summarize information from different data, taking into account opportunistic screening. Selection bias may affect the estimate, although it should be moderate using general population reference and may be partly compensated due to opportunistic screening. Conservative choices were systematically adopted, in order to avoid favourable estimates. This method provided estimates for the effect of organized biennial screening in France in the early nineteen-nineties. Survival analysis of screened cancer needs to be strengthened and results should be updated with longer follow-up and additional data.

Conflict of interest

None.

Appendix A

This appendix presents calculation of the relative risk of mortality between a screened population and a reference population. It is based on predicted number of cases from the Markov model and results of relative survival analyses (screened cancers and reference population). We denote j the node status ($j = 1, 2$ without or with node involvement respectively). And, for node status j, we denote d^j_{do}, d^j_{ki} and d^j_{ref} the predicted number of deaths among respectively screened cancers, interval cancers and reference population, S^j_d and S^j_{ref} the relative survival of screened cancers and cancers from the reference population. Then:

$$d^j_{do} = \sum_{i=1}^{8} n^d_{i,j}(1 - S^j_d(8 - i + 1)), \quad (1)$$

where $n^d_{i,j}$ is the number of screened cancers with node status j predicted for the year i

$$d^j_{ki} = \sum_{i=1}^{8} n^k_{i,j}(1 - S^j_{ref}(8 - i + 0.5)), \quad (2)$$

where $n^k_{i,j}$ is the number of interval cancers with node status j predicted for the year i

$$d^j_{ref} = \sum_{i=1}^{8} n^a_j(1 - S^j_{ref}(8 - i + 0.5)), \quad (3)$$

where n^a_j is the annual number of cancers with node status j predicted in the reference population.

The relative risk of mortality (RR), comparing predicted mortality in a screened and reference population, is estimated as follow (person-years are not reported since they are identical in numerator and denominator):

$$RR = \frac{d^1_{do} + d^2_{do} + d^1_{ki} + d^2_{ki}}{d^1_{ref} + d^2_{ref}} \quad (4)$$

References

[1] IARC handbooks of cancer prevention. vol. 7. Breast cancer screening. Lyon: IARC Press; 2002.
[2] Kerlikowske K, Rubin SM, Sandrock C, Ernster VL. Efficacy of screening mammography. A meta-analysis. JAMA 1995;273:149–54.
[3] Nyström L, Andersson I, Bjurstam N, Frisell J, Nordenskjöld B, Rutqvist LE. Long-term effects of mammography screening: updated overview of the Swedish randomised trials. Lancet 2002;359:909–19.
[4] Olsen O, Gotzsche PC. Cochrane review on screening for breast cancer with mammography. Lancet 2001;358:1340–2.
[5] Sarkeala T, Heinävaara S, Anttila A. Organised mammography screening reduces breast cancer mortality: a cohort studyfrom Finland. Int J Cancer 2008;122(3):614–9.
[6] Jonsson H, Bordás P, Wallin H, Nyström L, Lenner P. Service screening with mammography in Northern Sweden: effects on breast cancer mortality—an update. J Med Screen 2007;14(2):87–93.
[7] Swedish Organised Service Screening Evaluation Group. Reduction in breast cancer mortality from organized service screening with mammography: 1. Further confirmation with extended data. Cancer Epidemiol Biomarkers Prev 2006;15(1):45–51.
[8] Swedish Organised Service Screening Evaluation Group. Reduction in breast cancer mortality from the organised service screening with mammography: 2. Validation with alternative analytic methods. Cancer Epidemiol Biomarkers Prev 2006;15(1):52–6.
[9] Berry DA, Cronin KA, Plevritis SK, Fryback DG, Clarke L, Zelen M, et al. Effect of screening and adjuvant therapy on mortality from breast cancer. N Engl J Med 2005;353(17):1784–92.
[10] van Oortmarssen GJ, Habbema JD, van der Maas PJ, de Koning HJ, Collette HJ, Verbeek AL, et al. A model for breast cancer screening. Cancer 1990;66(7):1601–12.
[11] Tan T, van Oortmarssen GJ, de Koning HJ, Boer R, Habbema JD. The MISCAN-Fadia continuous tumor growth model for breast cancer. J Natl Cancer Inst Monogr 2006;36:56–65.
[12] Arveux P, Wait S, Schaffer P. Building a model to determine the cost-effectiveness of breast cancer screening in France. Eur J Cancer Care 2003;12(2):143–53.
[13] 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. J Epidemiol Biostat 1997;2:9–23.

[14] 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. J Epidemiol Biostat 1997;2:25–35.
[15] Allgood PC, Warwick J, Warren RM, Day NE, Duffy SW. A case–control study of the impact of the East Anglian breast screening programme on breast cancer mortality. Br J Cancer 2008;98(1):206–9.
[16] Puliti D, Miccinesi G, Collina N, De Lisi V, Federico M, Ferretti S, et al. Effectiveness of service screening: a case–control study to assess breast cancer mortality reduction. Br J Cancer 2008;99(3):423–7.
[17] Duffy SW, Hill C, Estève J. Quantitative methods for the evaluation of cancer screening. New York: Oxford University Press, 2001.
[18] Rue M, Vilaprinyo E, Lee S, Carles MD, Marcos-Gragera R, Pla R, et al. Effectiveness of early detection on breast cancer mortality in Catalonia (Spain). BMC Cancer 2009;9:326.
[19] Uhry Z, Hédelin G, Colonna M, Asselain B, Arveux P, Rogel A, et al. Multistate markov model in cancer screening evaluation: a brief review and a case study. Stat Methods Med Res 2010. doi: 10.1177/0962280209359848.
[20] Esteve J, Benhamou E, Croasdale M, Raymond L. Relative survival and the estimation of net survival: elements for further discussion. Stat Med 1990;9(5):529–38.
[21] Dickman PW, Sloggett A, Hills M, Hakulinen T. Regression models for relative survival. Stat Med 2004;23(1):51–64.
[22] Robert-Bobée I, Monteil C. Inégalités sociales de mortalité chez la femme (2005), « Quelles évolutions des différentiels sociaux de mortalité pour les femmes et les hommes? Tables de mortalité par catégorie sociale en 1975, 1982 et 1990 et Indices Standardisés de Mortalité par catégorie sociale en 1975, 1982, 1990 et 1999 », Document de travail, n° F0506, Insee.
[23] Day NE. Overdiagnosis and breast cancer screening. Breast Cancer Res 2005;7(5):228–9.
[24] 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. Stat Med 1995;14(14):1531–43.
[25] 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.
[26] 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). Int J Epidemiol 1991;20(4):852–8.
[27] 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. J Epidemiol Community Health 1996;50(1):68–71.
[28] Foucher Y, Giral M, Soulillou JP, Daures JP. A flexible semi-Markov model for interval-censored data and goodness-of-fit testing. Stat Methods Med Res 2008. doi: 10.1177/0962280208093889.
[29] Weedon-Fekjaer H, Lindqvist BH, Vatten LJ, Aalen OO, Tretli S. Breast cancer tumor growth estimated through mammography screening data. Breast Cancer Res 2008;10(3):R41.
[30] Duffy SW, Cuzick J, Tabar L, Vitak B, Chen HH, Yen MF, et al. Correcting for non-compliance bias in case–control studies to evaluate cancer screening programmes. Appl Stat 2002;51(2):235–43.
[31] Duport N, Ancelle-Park R, Boussac-Zarebska M, Uhry Z, Bloch J. Are breast cancer screening practices associated with sociodemographic status and healthcare access? Analysis of a French cross-sectional study. Eur J Cancer Prev 2008;17(3):218–24.
[32] Biesheuvel C, Barrat A, Howard K, Houssami N, Irwiq L. Effects of study methods and biases on estimates of invasive breast cancer overdetection with mammography screening: a systematic review. Lancet Oncol 2007;8(12):1129–37.
[33] Hery C, Ferlay J, Boniol M, Autier P. Changes in breast cancer incidence and mortality in middle-aged and elderly women in 28 countries with Caucasian majority populations. Ann Oncol 2009;19:1009–18.
[34] Duffy SW, Lynge E, Jonsson H, Ayyaz S, Olsen AH. Complexities in the estimation of overdiagnosis in breast cancer screening. Br J Cancer 2008;99(7):1176–8.
[35] Ravdin PM, Cronin KA, Howlader N, Berg CD, Chlebowski RT, Feuer EJ, et al. The decrease in breast-cancer incidence in 2003 in the United States. N Engl J Med 2007;356:1670–4.
[36] Parkin DM. Is the recent fall in incidence of post-menopausal breast cancer in UK related to changes in use of hormone replacement therapy? Eur J Cancer 2009;45(9):1649–53.
[37] Séradour B, Allemand H, Weill A, Ricordeau P. Changes by age in breast cancer incidence, mammography screening and hormone therapy use in France from 2000 to 2006. Bull Cancer 2009;96(4):E1–6.
[38] Paci E, Miccinesi G, Puliti D, Baldazzi P, De Lisi V, Falcini F, et al. Estimate of overdiagnosis of breast cancer due to mammography after adjustment for lead time. A service screening study in Italy. Breast Cancer Res 2006;8(6):R68.
[39] de Koning HJ, Draisma G, Fracheboud J, de Bruijn A. Overdiagnosis and overtreatment of breast cancer: microsimulation modelling estimates based onobserved screen and clinical data. Breast Cancer Res 2006;8(1):202.
[40] Morrell S, Barratt A, Irwig L, Howard K, Biesheuvel C, Armstrong B. Estimates of overdiagnosis of invasive breast cancer associated with screening mammography. Cancer Causes Control 2010;21(2):275–82.
[41] Jørgensen KJ, Gøtzsche PC. Overdiagnosis in publicly organised mammography screening programmes: systematic review of incidence trends. BMJ 2009;339:b2587.
[42] Jonsson H, Johansson R, Lenner P. Increased incidence of invasive breast cancer after the introduction of service screening with mammography in Sweden. Int J Cancer 2005;117(5):842–7.
[43] Seigneurin A, Exbrayat C, Labarère J, Colonna M. Comparison of interval breast cancer rates for two-versus single-view screening mammography: a population-based study. Breast 2009;18(5):284–8.
[44] Schopper D, de Wolf C. How effective are breast cancer screening programmes by mammography? Review of the current evidence. Eur J Cancer 2009;45(11):1916–23.
[45] de Gelder R, Bulliard JL, de Wolf C, et al. Cost-effectiveness of opportunistic versus organised mammography screening in Switzerland. Eur J Cancer 2009;45(1):127–38.