SORT 38 (2) July-December 2014, 139-160

Assessing the impact of early detection biases on breast cancer survival of Catalan women

Albert Roso-Llorach[1, 2], Carles Forné[3, 4], Francesc Macià[5, 6], Jaume Galceran[7], Rafael Marcos-Gragera[8] and Montserrat Rué[6, 9, *]

Abstract

Survival estimates for women with screen-detected breast cancer are affected by biases specific to early detection. Lead-time bias occurs due to the advance of diagnosis, and length-sampling bias because tumors detected on screening exams are more likely to have slower growth than tumors symptomatically detected. Methods proposed in the literature and simulation were used to assess the impact of these biases. If lead-time and length-sampling biases were not taken into account, the median survival time of screen-detected breast cancer cases may be overestimated by 5 years and the 5-year cumulative survival probability by between 2.5 to 5 percent units.

MSC: 62N02; 62P10.

Keywords: Breast cancer, early detection, screening, lead time bias, length bias, survival.

1. Introduction

Some types of cancer can be detected before they cause symptoms. The primary goal of cancer screening programs is to reduce mortality. Screening tests, such as mam-

* Corresponding author e-mail: montse.rue@cmb.udl.cat

[1] Unitat de Suport a la Recerca Terres de l'Ebre, Institut Universitari d'Investigació en Atenció Primària Jordi Gol (IDIAP Jordi Gol), Tortosa, Spain.

[2] Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Spain.

[3] Dept. Ciències Mèdiques Bàsiques, IRBLLEIDA-Universitat de Lleida, Spain.

[4] Oblikue Consulting, Barcelona, Spain.

[5] Servei d'Epidemiologia i Avaluació, Hospital del Mar-Parc de Salut Mar, Barcelona, Spain.

[6] Red de Investigación en Servicios de Salud y Enfermedades Crónicas (REDISSEC).

[7] Registre de Càncer de Tarragona, Fundació Lliga per a la Investigació i Prevenció del Càncer, Reus, IISPV, RTICC, Spain.

[8] Unitat d'Epidemiologia i Registre de Càncer de Girona (UERCG), Pla Director d'Oncologia. Departament de Salut. Institut d'Investigació Biomèdica de Girona (IdIBGi), Spain.

[9] Dept. Ciències Mèdiques Bàsiques, IRBLLEIDA-Universitat de Lleida, Avgda. Rovira Roure 80, 25198-Lleida, Spain.

Received: September 2013

Accepted: April 2014

mography, can detect cancer at an earlier stage compared to symptomatic diagnosis. It is expected that an early diagnosis will be associated with a better prognosis and consequently, with an increase of survival time. However, measuring the benefit of early detection as survival time from the date of diagnosis is confounded by two screening-specific biases: lead-time and length-sampling biases (Zelen and Feinleib, 1969).

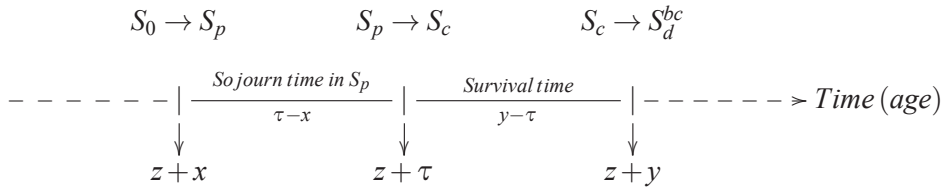
For a screen-detected cancer, the lead-time is defined as the time gained by diagnosing the disease before the patient experiences symptoms. Even if early diagnosis and early treatment had no benefit, the survival of early detected cancer cases would be longer than the survival of clinical cases (see Figure A.1 in the Appendix). Length-sampling bias arises because screen-detected cancers are more likely to have slower growth than non-screen detected cancers. It seems reasonable to assume that the clinical course of the disease is positively correlated with its pre-clinical course. Thus, patients with screen-detected cancers survive longer in part because their tumors are less aggressive. Therefore the difference in survival cannot be only attributed to the early detection (see Figure A.1 in the Appendix). Different authors have studied the effect of these biases in the survival functions of women with screen-detected breast cancer (BC) and have proposed several corrections (Walter and Stitt, 1987; Xu and Prorok, 1995; Xu, Fagerstrom and Prorok, 1999; Duffy et al., 2008 and Mahnken et al., 2008). The goals of this study are: 1) To review the methods of bias correction for BC; 2) To obtain bias-corrected survival estimates of the screen-detected cases; and 3) To evaluate the impact of the lead-time and length-sampling biases. The rest of the paper is organized as follows. Section 2 reviews the existing methods in the literature for bias correction and describes the statistical methods used, including a simulation study. Section 3 presents the results, and Section 4 is a general discussion.

2. Methods

2.1. Breast cancer early detection model

As defined by Zelen and Feinleib (1969), the progress of BC can be characterized as a stochastic process, assuming that each individual in a specific population is in one of these three states: disease-free (S_0); the pre-clinical or asymptomatic state (S_p), when the disease can be diagnosed by a special exam; and the clinical or symptomatic state (S_c). Sometimes an absorbing state (S_d^{bc}) referring to death from BC can be added.

Based on this early work, Lee and Zelen (LZ) proposed a stochastic model for predicting the mortality of the early detection programs as a function of the characteristics of the early detection scenario (Lee and Zelen, 1998, 2008). The assumptions of the LZ model are: (1) progressive disease; (2) age-dependent transitions into the different states, $S_0 \rightarrow S_p \rightarrow S_c \rightarrow S_d^{bc}$; (3) age-dependent examination sensitivity; (4) age-dependent sojourn times in each state; and (5) exam-diagnosed cases have a stage-shift in the direction of more favorable prognosis relative to the distribution of stages in symptomatic detection.



Note that the transition $S_0 \to S_p$ is never observed and the transition $S_p \to S_c$ refers to the disease incidence. If the early detection exam does diagnose the disease in the pre-clinical state, the transition $S_p \to S_c$ will never be observed.

The LZ model considers:

- n screening exams at times $t_0 < t_1 < ... < t_{n-1}$. It is assumed that $t_0 = 0$ and $z = age$ at t_0.
- Three chronological times (see above schema):
  - x: time at entering S_p, $z+x$: age when entering S_p. The time x is not observed but can be derived from the incidence function and the distribution of sojourn time in the S_p state. x takes a negative value if the transition to S_p occurs before the age at first exam, z.
  - τ: time at entering S_c, $z+\tau$: age at entering S_c. The time τ can not be observed in cases detected by exam, only in the clinically detected cases. For cases detected by exam, τ can be estimated.
  - y: time at death, $z+y$: age at death. Then $x < \tau < y$
- Sojourn time in S_p: $\tau - x$
- Sojourn time in S_c: $y - \tau$

The LZ basic model calculates the cumulative probability of death for the cohort group exposed to any screening program after T years of follow-up. Similarly, the cumulative probability of death for the cohort group not exposed to screening can be calculated. These probabilities are used to calculate the possible reduction in mortality from an early detection program after T years of follow-up and can be obtained as follows.

Survival distributions for exam-diagnosed, interval, and control cases are assumed to be conditional on the stage at diagnosis and treatment, but are not dependent on the mode of diagnosis. The LZ model assumes k disease stages which describe the severity of a person's cancer based on the size and/or extent of the tumor. If $\phi_s(j)$, $\phi_i(j)$ and $\phi_c(j)$ represent the probability of being diagnosed at stage j, $j = 1,\ldots,k$ for exam-diagnosed, interval and control cases, respectively, and $f_j(t|z+\tau)$ is the probability density function (pdf) of survival time t among subjects who would have been clinically diagnosed at stage j in the absence of screening, then the survival time *pdf*s of the exam-diagnosed, interval and control cases are the mixtures:

$$g_s(t|z+\tau) = \sum_{j=1}^{k} \phi_s(j) f_j(t|z+\tau), \quad g_i(t|z+\tau) = \sum_{j=1}^{k} \phi_i(j) f_j(t|z+\tau)$$

and

$$g_c(t|z+\tau) = \sum_{j=1}^{k} \phi_c(j) f_j(t|z+\tau),$$

respectively. In other words, the g density functions are obtained by weighting the f functions by the distribution of disease stages at diagnosis. Since screening will appear to increase survival time, the LZ model controls for *lead-time* bias by setting the origin of survival time for the screened, interval, and clinical cases at the time of clinical diagnosis. Consequently, there is an implied *guarantee time* for disease-specific survival, that is, the cases diagnosed earlier would have been alive at the time the disease would have been clinically diagnosed. This guarantee time, also called *lead-time*, is a random variable and is incorporated into the equations of the model. Explicitly, the lead-time is $\tau - t_r$ where τ is the time at which the individual enters the clinical state and t_r is the time at which the r detection exam, when the disease is diagnosed, is given.

2.2. Methods for correcting the biases specific to early detection

After reviewing the literature, we selected the methods of Walter and Stitt (1987), Xu and Prorok (1995), Xu et al. (1999) and Duffy et al. (2008). All these authors assume the progressive disease model aforementioned with an exponential distribution of the sojourn time in the pre-clinical state. The observed survival time, Z, after diagnosis by screening is defined as $Z = X + Y$, Y is the lead-time, and X the post-lead survival time (the time from clinical detection to death or the end of study). X is the time of interest, free of biases.

2.2.1. The Walter and Stitt method

Walter and Stitt (1987) developed a model for the survival of screen-detected cases, with a hazard function that depends on an individual's lead-time, Y, the duration of the sojourn time in the pre-clinical state and the time since diagnosis, Z. Their main assumptions were that the hazard function considers a guarantee time from the screening detection until when the disease would become clinical and an exponential distribution for the lead-time, Y (Walter and Day, 1983). The authors showed that if the post-lead-time, X, can be assumed to have an exponential distribution, the corresponding parameter can be estimated by maximum likelihood using life-table methods.

2.2.2. The Xu and Prorok method

Xu and Prorok (1995) developed a model under the assumption of an exponential distribution for the lead-time and independence between the lead-time and post-lead-time. They presented a method to estimate the survival function of the post-lead time, X, of screen-detected cancer cases based on the observed total survival time, Z. The authors relaxed the parametric assumption for the post-lead-time and obtained the non-parametric maximum likelihood estimator (NPMLE) of the survival function of the post-lead time, X.

2.2.3. The Xu et al. method

As Xu and Prorok mentioned, it seems biologically reasonable that the lead-time and the post-lead-time are positively correlated. Xu et al. (1999) introduced a new model that involved dependence between the lead- and post-lead-time through nuisance variables to ensure positive correlation. Several levels of correlation were studied. They applied the Xu and Prorok method on the new model to obtain the NPMLE of the post-lead-time survival function.

2.2.4. The Duffy et al. method

Duffy et al. (2008) proposed a simple correction for lead time, assuming an exponential distribution of the sojourn time in the pre-clinical state. The additional follow-up due to lead-time is estimated individually for each patient with a screen-detected cancer as the expected lead-time conditional on its being less than the observed survival time or time to last follow-up. The expression of the expected lead-time depends on whether the patient died of BC or not. The corrected survival time, for screen-detected cases, is obtained subtracting the expected lead-time from the observed survival time.

2.3. Data

BC survival data were obtained from the Girona and Tarragona population-based cancer registries (PCR) in Catalonia (both provinces representing 20% of the total Catalan population and covering either urban or rural areas). Data from Girona were provided directly by the Girona Cancer Registry and data from Tarragona was obtained through the Foundation League for the Research and Prevention of Cancer (FUNCA). Given that the BC incidence and mortality rates in the Girona and Tarragona registries were similar, both datasets were merged. The PCR sample included 1,221 women residing in the province of Girona and diagnosed between 2002 and 2006, and 2,149 women residing in the province of Tarragona and diagnosed between 2000 and 2005.

We also obtained BC survival data from the hospital cancer registry of Parc de Salut Mar (HCR-PSMAR) in the city of Barcelona. The HCR-PSMAR included BC tumours from women attending an early detection program (screen-detected or not) and also

BC tumours from other women living in the hospital area. The HCR-PSMAR sample included 1,704 women diagnosed with BC between 1996 and 2006. BC cases in this study refer to invasive BC. Ductal carcinoma in situ (DCIS) cases were not included.

2.4. Statistical analysis

2.4.1. Survival analysis

First, we estimated the biased BC specific survival using the Kaplan-Meier method, assuming that BC was the single cause of death. We considered death from BC as the event of interest. Deaths from other causes (OC) or lost to follow-up (either dropouts or withdrawals) were treated as right-censored observations. Censoring was assumed to be non-informative. Survival time was calculated as the difference between the date of diagnosis and the minimum of time to the event and censored time. Then, we applied the methods described in Section 2.2 in order to correct the BC-specific survival of screen-detected cases. We assumed an exponential distribution with scale parameter 0.25 for the lead-time. This assumption was based on the values proposed by Lee and Zelen (2006) for the mean sojourn time in the pre-clinical state, the previous work of Zelen and Feinleib (1969), the age at diagnosis distribution of the studied cases, and the simulation study described in 2.4.2. For the method of Xu et al. (1999), we considered a dependence parameter 0.5 corresponding to a moderate dependence between lead-time and post-lead-time. All analyses were performed with R version 3.0.1 (R Core Team, 2013).

2.4.2. Simulation study

Since the observed data were characterized by heavy right censoring, we conducted a simulation study. The main goal of the simulation study was to estimate the lead-time and length biases under different screening strategies and to compare the results with those obtained using the correction methods described in Section 2.2. The simulation reproduces the individual life histories of women initially in the disease-free state. Our simulation model considered the Lee and Zelen model inputs for Catalonia (Vilaprinyo et al., 2008, 2009; Martinez-Alonso et al., 2010) and additional assumptions described below. For simplicity, in the following sections t refers to chronological time or age.

Initial parameters We used observed or predicted data on BC incidence and mortality for the cohort of Catalan women born in 1950. We assumed a sample size of $n = 100,000$ women. The time horizon was 0-85 years of age, we only considered BC incident cases before age 85 and stopped the follow-up at age 85. We grouped the data by age, considering J yearly disjoint intervals $(a_{j-1}, a_j]$ for $j = 1, \ldots, J$, where $a_0 = 0$. We assumed the values proposed by Lee and Zelen (2006) for the age-dependent examination sensitivity, $\beta(t)$, and the exponential distribution with age-dependent mean, $m(t)$, for sojourn time in S_p. The $m(t)$ in years was: 2 for women 40 years old or younger,

4 for women older than 50 years and the linear interpolation $m(t) = -6 + 0.2 * age$ for women aged 40-50 years. The periodicity of the exams was annual or biennial. The initial ages of screening schedules were 40 and 50 years, while the ages at the last examination were 68 years for biennial and 69 years for annual strategies, resulting in four screening strategies. Bivariate correlated data of sojourn times in S_p and S_c were simulated using copula models (Trivedi and Zimmer, 2007). We chose the Clayton's Archimedean copula because it has some interesting features. For example, it is adequate for positive associations between times. Under the Clayton's copula model, three different dependence parameters were chosen, $\alpha \in \{1, 5/4, 3/2\}$; they represent values for Kendall's tau of $\tau_K \in \{0, 1/9, 1/5\}$ ranging from no association to moderate association.

Death from causes other than breast cancer The age-specific death rates from OC for Catalan women, by birth cohort, were used as the hazard function in a survival process where failure was death from OC. Then, ages at death from OC were sampled using the inverse transformation of the cumulative survival function.

Generation of the pre-clinical cases We used Catalan BC incidence rates, estimated assuming no screening for BC (Martinez-Alonso et al., 2010), to obtain the transition probabilities to the pre-clinical state using the method described by Lee and Zelen (1998). We considered these transition probabilities as the hazard in a survival process, where failure consists of entering S_p. Using the same reasoning as for OC, an age when entering S_p was generated for each simulated woman.

Generation of the age at entering the clinical state S_c Some authors have provided evidence that the sojourn time in the pre-clinical state is exponentially distributed (Zelen and Feinleib, 1969; Walter and Day, 1983). A sojourn time in S_p was sampled assuming an age-dependent exponential distribution with mean $m(t)$. Then an age when entering S_c was generated adding the sojourn time to the age at entering S_p, for each simulated woman that transitioned to S_p.

Generation of the screen-detected and the interval cancer cases For women that entered S_p, we considered that their BC could be screen-detected if they received screening exams during their sojourn time in S_p. To decide whether the result of an exam was positive or negative we used a Bernoulli random variable with success probability the sensitivity of the exam, $\beta(t)$. The cases diagnosed at the interval between two exams were considered as interval cases.

Death from breast cancer We used the Clayton's copula, as described in Trivedi and Zimmer (2007), to generate a survival time from the BC diagnosis, using the Catalan age-specific survival functions for BC (Vilaprinyo et al., 2009). The survival time was correlated with the sojourn time in S_p through the copula function.

For screen-detected cases, we considered two assumptions for the survival time: with and without benefit of early detection. When survival benefit was assumed, the survival *pdf*s for screen-detected, interval, and clinical cases were obtained weighting the age- and stage-specific survival *pdf*s by the distribution of disease stages at diagnosis. (See Section 2.1 for more details). The distribution of disease stages at diagnosis for screen-detected, interval and clinical BC cases is shown in Table A.1 in the Appendix.

The no-survival benefit assumption was based on a systematic review that reported a non-statistically significant reduction in BC mortality for trials with adequate randomization (Gotzsche and Nielsen, 2009). When no-survival benefit from screening was assumed, we used the clinical stages distribution for screen-detected, interval and clinical cases.

Once the survival time was generated, the age of death from BC was obtained adding the survival time to the age when entering the clinical state S_c for the screened, interval and clinical cases. In that way, there is no lead-time bias for the screen-detected cases.

Age at death We obtained the age of death as the minimum between age at BC death and age at OC death, assuming that both events were independent. A total of 24 scenarios were analysed considering the two assumptions for the survival benefit of early detection, the four screening strategies and the three copula parameters.

The simulation code was developed in R version 3.0.1 (R Core Team, 2013). For each scenario, to generate one dataset, the algorithm ran for approximately 45 seconds on a MacBook Pro machine with 2.4 Ghz Intel Core 2 Duo processor with 4 GB of RAM memory. For each scenario $B = 100$ datasets were generated.

2.4.3. Estimation of the lead-time and length-sampling biases for screen-detected cases

The lead-times for the screen-detected cases were obtained as the difference between the age at entering the clinical state and the age at detection. To estimate the mean lead-time of each scenario, first we obtained the mean lead-time within each dataset and then we calculated the mean and the empirical standard error of the 100 dataset means.

To estimate the length bias, first we obtained the median survival time of screen detected cases corrected by the lead-time bias. Then we obtained the median survival time of the background scenario (no screening). Finally, the difference of the two median survival times was considered the length bias effect on the median survival time of screen-detected cases. For the scenarios with no benefit of screening and independence between sojourn time in the pre-clinical state and survival time, the expected length bias would be zero.

2.4.4. Comparison of the methods of bias correction

We calculated the root mean square error (RMSE) between the simulated unbiased cumulative survival and the corrected cumulative survival for the different methods of bias correction. The RMSE gives the standard deviation of the model prediction errors. A smaller value indicates a better model performance. To compute the RMSE we considered the first 25 years of follow-up. The mean RMSE over the 100 simulations was obtained for each scenario (Burton et al., 2006).

2.4.5. Validation

We have compared our results with results in the literature on cumulative incidence and BC cumulative survival. In addition, we have compared a) the frequencies of screen-detected and interval cancer, by age-group; and b) the sensitivity of the program, with the results of the INterval CAncer (INCA) study in Spain, which included 645,764 women aged 45/50 to 69 years that participated biennially in seven population-based screening programs, from January 2000 to December 2006 (Blanch et al., 2014 and Domingo et al., 2014). The cohort was followed until June 2009 for breast cancer identification, resulting in 5,309 cases screen-diagnosed and 1,653 interval cancers. The sensitivity of the program was defined as the ratio of the number of tumors detected in the screening exams between all the detected tumors.

3. Results

3.1. Observed and corrected cumulative survival. Data from the cancer registries

Table 1 presents the median follow-up time and the censoring percentage for screen-detected and clinical cases, according to BC survival status. Both the PCR and HCR-PSMAR samples presented a large percentage of right censoring, which was around 95% or higher for screen-detected cases. The median follow-up time was shorter for the PCR sample.

Table 1: *Follow-up time and survival status for the two studied samples.*

	Population Cancer Registries Girona and Tarragona		Hospital Cancer Registry PSMAR	
	No BC death	BC death	No BC death	BC death
Screen-detected cases (*n*)	633	19	463	27
Median of follow-up (years)	5.46	3.82	7.19	4.31
Percent (%)	97.1	2.9	94.5	5.5
Clinical cases (*n*)	2284	434	988	226
Median of follow-up (years)	5.10	2.31	6.43	3.10
Percent (%)	84.0	16.0	81.4	18.6

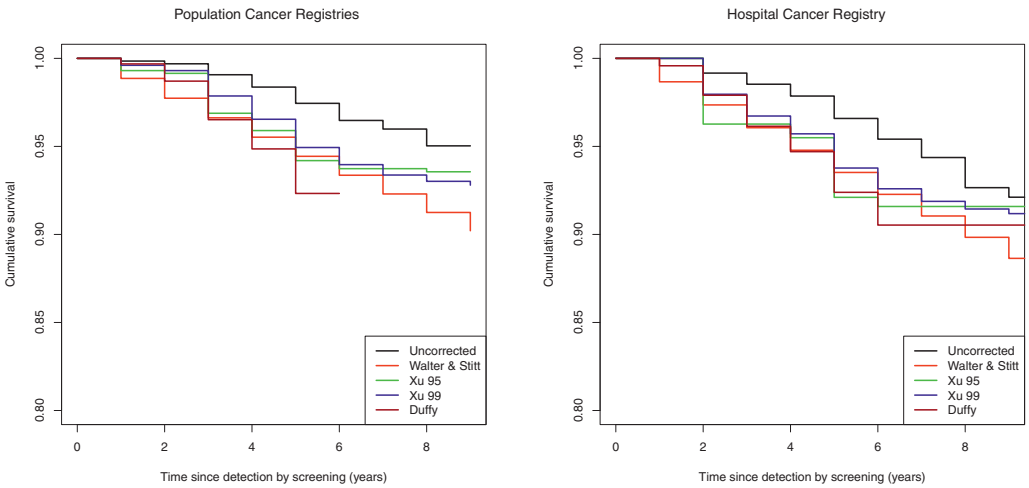


Figure 1: *Observed (black) and corrected survival (colours) for each method of correction, for screen-detected cases.*

Figure 1 shows the observed and corrected BC survival of screen-detected cases, using the methods described in Section 2.2, for both studied samples. The corrected cumulative survival curves grouped together below the observed survival curve. Table 2 presents the observed and corrected cumulative survival rates at five years after BC detection. Differences of observed and corrected cumulative survival varied from 2.5 to 5.1% units. Observed cumulative survival rate at 5 years around 97% decreased to 94 or 92% after correction. The higher difference was observed for the Duffy method in the PCR sample (5.1%) followed by the Xu and Prorok (4.5%) and the Duffy methods in the HCR-PSMAR sample (4.2%).

Table 2: *Observed and corrected survival rates at five years after breast cancer detection.*

Cumulative Survival	Population Cancer Registries Girona and Tarragona	Hospital Cancer Registry PSMAR
Observed (uncorrected)	97.44	96.59
Walter and Stitt	94.44	93.52
Xu and Prorok	94.19	92.11
Xu et al.	94.94	93.77
Duffy et al.	92.33	92.39

3.2. Simulation study

The detailed simulation results for all the 24 scenarios can be found in the Appendix (Tables A.2, A.3, A.4 and A.5 and Figure A.2).

Table 3 describes the lead-time (mean and standard error of the 100 simulated datasets for each of the 24 scenarios), overall and stratified by age at entering S_p, for

Table 3: *Estimated lead-time (years) for screen-detected cases, overall and by age at entering the pre-clinical state. Mean and standard error (S.E.) of the 100 simulated datasets for each screening strategy.*

	Age at entering the pre-clinical state							
	Overall		< 40 yrs		$40-49$ yrs		≥ 50 yrs	
Strategy	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
A4069	3.67	0.06	1.98	0.26	3.17	0.14	3.83	0.07
B4068	3.69	0.07	1.98	0.27	3.22	0.15	3.85	0.08
A5069	3.81	0.06	1.73	1.60	3.56	0.22	3.84	0.07
B5068	3.81	0.07	1.63	1.59	3.56	0.22	3.84	0.08

A4069: Annual exams in the age interval 40-69 years. B4068: Biennial exams in the age interval 40-68 years.
A5069: Annual exams in the age interval 50-69 years. B5068: Biennial exams in the age interval 50-68 years.

the four screening strategies. Mean lead-times for all the strategies, by age group, were similar, with an increasing trend by age at entering S_p. It is important to notice that the mean lead-times correspond to screen-detected cancers only.

Figure 2 shows the cumulative observed (solid) and corrected (dashed) BC survival after diagnosis of BC for screen-detected cases, for biennial screening strategies. The figure corresponds to one of the 100 simulated datasets for $\alpha = 1.25$ with (left) and without (right) survival benefit. The separation of the curves is more marked in the assumption of no survival benefit, mainly for the 5 to 10 years follow-up time interval.

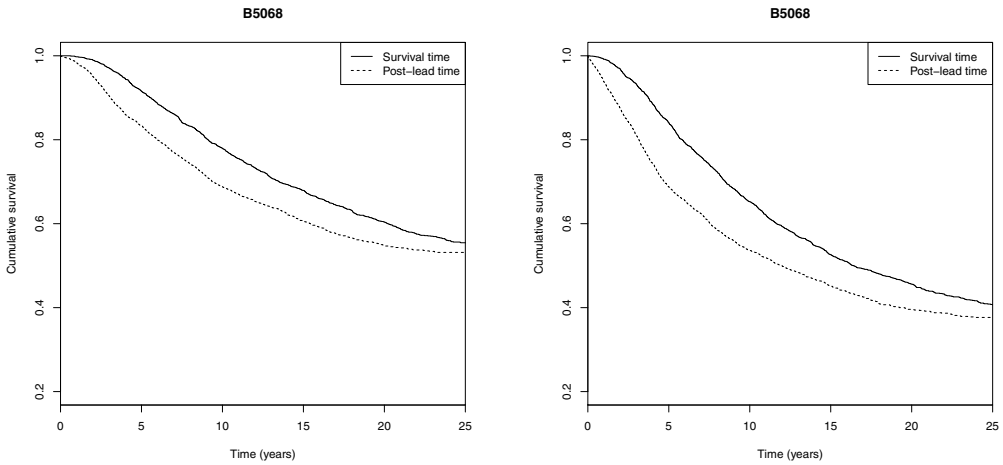


Figure 2: *BC cause-specific survival of screen-detected cases. B5068: Biennial exams in the age interval 50-68 years.* $\alpha = 1.25$*, left: with screening benefit, right: without screening benefit.*

Table 4 presents the mean and standard error estimates of the median survival time and the median post-lead-time for the screen-detected cases with the assumption of no benefit. The lead-time and length biases are also summarized. For each screening strategy, both the survival time and post-lead-time increase as α increases. This result is

Table 4: *Median total survival time (biased), median post-lead-time (corrected) and early detection biases for screen-detected cases, without benefit of screening. Different assumptions (values of α) of correlation between time in S_p and survival time.*

	Without benefit of screening							
$\alpha = 1$	Median survival time		Median post-lead time		Median lead-time bias		Median length bias	
Strategy	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
A4069	15.83	0.43	11.20	0.44	4.63	0.23	-0.03	0.31
B4068	15.83	0.48	11.20	0.49	4.63	0.26	-0.03	0.35
A5069	15.69	0.43	10.92	0.42	4.76	0.21	-0.31	0.31
B5068	15.66	0.48	10.92	0.46	4.75	0.27	-0.31	0.38
$\alpha = 1.25$	Median survival time		Median post-lead time		Median lead-time bias		Median length bias	
Strategy	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
A4069	16.09	0.52	11.74	0.45	4.35	0.23	1.11	0.34
B4068	16.69	0.59	12.32	0.53	4.38	0.25	1.69	0.41
A5069	15.99	0.48	11.50	0.45	4.49	0.23	0.87	0.34
B5068	16.49	0.59	12.01	0.53	4.48	0.24	1.38	0.44
$\alpha = 1.5$	Median survival time		Median post-lead time		Median lead-time bias		Median length bias	
Strategy	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
A4069	16.40	0.54	12.23	0.46	4.17	0.20	2.00	0.36
B4068	17.71	0.70	13.37	0.64	4.34	0.27	3.13	0.53
A5069	16.35	0.51	12.03	0.45	4.31	0.26	1.80	0.35
B5068	17.45	0.65	13.03	0.57	4.42	0.27	2.79	0.48

A4069: Annual exams in the age interval 40-69 years. B4068: Biennial exams in the age interval 40-68 years.
A5069: Annual exams in the age interval 50-69 years. B5068: Biennial exams in the age interval 50-68 years.

consistent with the facts: 1) screen-detected tumors have a longer sojourn time in S_p; and 2) higher values of α indicate higher correlation between time in S_p and survival time, therefore, longer sojourn times will have more chances of being followed by longer survival times and post-lead times. Median lead-time is higher than 4 years in all the screening strategies and decreases as α increases. In contrast, the median length bias is near zero for $\alpha = 1$ and increases with α. For $\alpha = 1.25$, which indicates moderate correlation between sojourn time in S_p and survival time, the median length bias takes values around 1 year. While the lead-time is similar in annual and biennial strategies, the length bias is higher in biennial than annual strategies.

Table 5 provides the RMSE mean between the simulated and predicted survival when the bias correction methods were used, for each screening scenario. For all scenarios, the Xu and Prorok and the Duffy et al. methods outperformed the other methods in terms of mean RMSE. The Walter and Stitt method obtained the worst mean RMSE in all

Table 5: *Root mean square error (RMSE)* $\times 10^{-2}$ *obtained using the bias correction methods described in section 2.2.*

With survival	$\alpha = 1$				$\alpha = 1.25$				$\alpha = 1.5$			
benefit	A4069	B4068	A5069	B5068	A4069	B4068	A5069	B5068	A4069	B4068	A5069	B5068
Walter and Stitt	4.44	4.75	3.47	3.88	4.28	4.53	3.29	3.62	4.13	4.28	3.15	3.39
Xu and Prorok	1.23	1.35	1.31	1.41	0.89	0.99	0.94	1.04	1.14	1.23	1.19	1.29
Xu et al.	2.27	2.44	2.44	2.61	1.75	1.85	1.80	1.90	1.63	1.68	1.57	1.64
Duffy et al.	1.52	1.68	1.61	1.78	1.02	1.10	1.18	1.25	0.92	0.94	1.12	1.13
Without survival	$\alpha = 1$				$\alpha = 1.25$				$\alpha = 1.5$			
benefit	A4069	B4068	A5069	B5068	A4069	B4068	A5069	B5068	A4069	B4068	A5069	B5068
Walter and Stitt	7.22	7.16	6.28	6.32	6.99	6.84	6.00	5.96	6.74	6.52	5.70	5.60
Xu and Prorok	1.89	1.93	1.95	1.99	1.21	1.26	1.27	1.31	1.46	1.48	1.56	1.57
Xu et al.	4.04	4.03	4.31	4.29	3.20	3.11	3.35	3.26	2.81	2.70	2.87	2.75
Duffy et al.	3.15	3.16	3.34	3.34	2.25	2.19	2.46	2.38	1.71	1.61	1.95	1.83

A4069: Annual exams in the age interval 40-69 years.

B4068: Biennial exams in the age interval 40-68 years.

A5069: Annual exams in the age interval 50-69 years.

B5068: Biennial exams in the age interval 50-68 years.

scenarios and the Xu and Prorok method performed better in scenarios with moderate association; on the other hand the Xu et al. method performed better in moderate or strong association scenarios and with survival benefit.

3.3. Validation

Our cumulative incidence estimate in the 0-85 age interval was 7.81% for the cohort of Catalan women born in 1950 (Table A.2 in the Appendix). The results are consistent with cross-sectional estimates in the 0-74 age-interval of 7.01% in 1995 and 7.89% in 2002, for Catalan women (Borras et al., 2008). Moreover, the Catalan survival rate at five years was 80.9 for women diagnosed with BC in the period 1995-1999 (Galceran et al., 2008). The corresponding estimate in our simulation study, assuming that there was a screening benefit, is somewhat lower, 76.1%.

Our simulated results show that around 40 to 50% of women diagnosed with BC are expected to die of the disease (Table A.5 in the Appendix). These results are comparable with those obtained by Bush et al. who reported that non-BC deaths accounted for almost half of deaths among BC patients in the 15 years following diagnosis (Bush et al., 2010).

Our simulated data estimated percentages of interval cases among all BC cases equal to 30.6% and 28.7% in the age groups 50-59 and 60-69 years, respectively, for the scenario B50-68. Corresponding data for the INCA study were 36% and 26%, respectively (data not published).

Our estimated overall program sensitivity for B50-68 was 70.5%. This value in the INCA study was 68.1% (data not published).

4. Discussion

4.1. Principal findings

This study used BC registry data and simulations to correct BC survival estimates and to assess the impact of lead-time and length sampling biases on survival estimates of screen-detected BC. When the observed survival estimates from the PCR or the HCR-PSMAR were corrected for lead-time bias, the cumulative survival estimates at 5 years decreased between 2.5 to 5.1 percent units, depending on the correction method used. The simulation results showed that, except the Walter and Stitt method, the other three methods for correcting biases performed without major differences. Furthermore, the most accurate correction for the survival estimate was obtained with one or another method depending on different settings. In addition, the simulation results also showed that: 1) screening for BC annually or biennially after 40 years of age brings the age at diagnosis for screen detected cancers forward by more than 3 years; 2) median survival time of screen-detected cases may be overestimated by more than 4 years due to lead-time bias; and 3) assuming a moderate correlation between sojourn time in the pre-

clinical state and survival time (parameter $\alpha = 1.25$), women with screen-detected BC may have a median survival time (already corrected by lead-time) around 1 year or more longer than non-screened women due to length bias. Overall, median survival of screen-detected cases might have been overestimated by 5 years if no corrections for these biases were made.

4.2. Comparison with other studies

Some authors, such as Kafadar and Prorok (2009), have assumed that the benefit of screening is zero to be able to estimate the length bias. According to Kafadar and Prorok, since survival time for screen-detected cases confounds the effects of lead-time, benefit time, and length-sampling bias, studies that use survival time to evaluate screening programs need to take account of these effects.

Shen et al. (2005) found an apparent survival benefit beyond stage shift for patients with screen-detected BC compared with patients with BC detected otherwise. They concluded that method of detection is an important prognostic factor for BC survival, even after adjusting for known tumor characteristics. This result is consistent with our results which indicate a non-negligible length bias effect.

Lehtimaki et al. (2011) performed a multivariate analysis to assess the effect of methods of detection on BC survival, adjusted by tumor size, node involvement, differentiation grade, hormonal status and ductal type. The method of detection was an independent prognostic factor, with a hazard ratio of 1.69 (95% confidence interval = 1.06 to 2.70) between patients whose tumors were detected outside screening and those whose tumors were screen-detected. The authors conclude that survival differences could not be explained completely by lead-time and length bias-related variables, although they may have not completely corrected these biases when adjusting by known risk factors.

4.3. Limitations

This study has several limitations. First, data from the PCR and the HCR-PSMAR presented a large percentage of right censoring, that hindered the application of the methods of bias correction and interpretation of the results. Our simulation study tried to overcome this limitation by extending the follow-up and therefore increasing the number of events. Second, our model relies on data and assumptions that may be not correct. For instance, a) the older age-specific BC incidence and mortality rates for the studied 1950 cohort were projected using an age-period-cohort model. b) The distribution of disease stages at diagnosis for annual or biennial strategies or for screen-detected, clinical or interval cases was taken from US data due to non-availability of annual screening data from the Catalan or Spanish registries. c) We assumed independence between death from BC and other causes. d) We could not test the appropriateness of the copula parameters that correlate both sojourn and survival times. Thus, we used several values compatible with low, medium or high correlation assumptions between the sojourn times. In any

case, many of the simulated results are consistent with the literature and the trends observed are compatible with the studied screening scenarios, therefore we think that our estimates of lead-time and length sampling biases are reliable.

4.4. Conclusion

Survival estimates of screen-detected BC cases are affected by the lead-time and length-sampling biases. The size of these biases depends on the starting age and periodicity of the screening exams. If lead-time and length-sampling bias were not taken into account, the median survival time of screen-detected BC cases may be overestimated by 5 years and the cumulative survival at 5 years may be overestimated between 2.5 to 5 percent. Our results illustrate the importance of correcting or controlling these biases when assessing the benefit of screening mammography. The Xu and Prorok, Duffy et al. and Xu et al. methods for correcting biases outperformed the Walter and Stitt method, with slight differences depending on the scenarios' assumptions.

Acknowledgements

In Memory of Professor Marvin Zelen, without whose enormous contributions to the field of statistics this article would not have been possible.

This study was funded by grant PS09/01340 from the Health Research Fund (Fondo de Investigación Sanitaria) of the Spanish Ministry of Health. We are also indebted to the INCA Study Group for providing data that was used to validate our simulation study and to the research group Grup de Recerca en Anàlisi Estadística de la Supervivència (GRASS) for fruitful discussions. We also thank the reviewers of this paper for their valuable comments on a preliminary version of the manuscript and JP Glutting for review and editing.

Appendix

Table A.1: *Distribution of stages at diagnosis of BC.*

Age (years)	Stages[1] I	II-	II+	III	IV
	Background[1,2]				
40-49	0.3008	0.2277	0.3091	0.0999	0.0625
50-59	0.2868	0.2176	0.3111	0.1021	0.0825
60-69	0.3028	0.2225	0.2713	0.0974	0.1061
70-79	0.3157	0.2671	0.2227	0.0983	0.0961

Table A.1 (cont): *Distribution of stages at diagnosis of BC.*

Annual screening. Screen-detected cases[1,3]					
40-49	0.6200	0.1131	0.2141	0.0436	0.0092
50-59	0.6669	0.1057	0.1935	0.0296	0.0043
60-69	0.7641	0.0739	0.1412	0.016	0.0047
70-79	0.7821	0.0875	0.1067	0.0165	0.0072
Annual screening. Interval cases[1,3]					
40-49	0.4644	0.1903	0.2598	0.0667	0.0188
50-59	0.4501	0.1744	0.2976	0.0665	0.0113
60-69	0.5417	0.1532	0.2320	0.0591	0.0141
70-79	0.5446	0.2345	0.1583	0.0496	0.013
Biennial screening. Screen-detected cases[1,3]					
40-49	0.5839	0.1217	0.2360	0.0438	0.0146
50-59	0.6210	0.1472	0.1734	0.0423	0.0161
60-69	0.6563	0.1295	0.1830	0.0246	0.0067
70-79	0.7287	0.1311	0.1128	0.0137	0.0137
Biennial screening. Interval cases[1,3]					
40-49	0.3673	0.2246	0.3099	0.0819	0.0164
50-59	0.2945	0.2609	0.2648	0.1166	0.0632
60-69	0.4077	0.2231	0.2672	0.0744	0.0275
70-79	0.4336	0.2885	0.1770	0.0673	0.0336

[1] American Joint Committee on Cancer (AJCC) stage distribution.

[2] From Surveillance, Epidemiology, and End Results (SEER).

[3] From Breast Cancer Surveillance Consortium (BCSC).

Table A.2: *Pre-clinical state summary and cumulative incidence. Background scenario. One hundred simulated scenarios for each screening strategy. Time horizon 0-85 years.*

Parameter	Mean	S.E.
Cumulative transition to S_p (%)	9.08	0.09
Cumulative incidence (%)	7.81	0.08
Mean sojourn time in S_p (years)	3.24	0.04
Mean sojourn time in $S_p \leq 40$ (years)	2.00	0.11
Mean sojourn time in S_p $40-50$ (years)	3.16	0.13
Mean sojourn time in $S_p > 50$ (years)	4.01	0.05

Table A.3:*Screening summary. One hundred simulated scenarios for each screening strategy. Time horizon 0-85 years.*

Strategy[1]	Mean sojourn time SD[2]		Mean sojourn time I[2]		Program sensitivity[3]		Cumulative incidence[3]	
	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
A4069	4.40	0.06	0.66	0.02	47.40	0.62	7.87	0.08
B4068	4.91	0.07	1.12	0.03	37.39	0.56	7.85	0.08
A5069	4.67	0.06	0.72	0.03	41.49	0.60	7.86	0.08
B5068	5.15	0.08	1.19	0.03	33.15	0.56	7.85	0.08

[1] A4069: Annual exams in the age interval 40-69 years.
B4068: Biennial exams in the age interval 40-68 years.
A5069 Annual exams in the age interval 50-69 years.
B5068: Biennial exams in the age interval 50-68 years.

[2] SD: Screen-detected cases, I: Interval cases.

[3] Expressed as percentage.

Table A.4: *Median survival summary for interval cancer cases. One hundred simulated scenarios for each screening strategy. Time horizon 0-85 years.*

	Survival time for interval cancer cases			
$\alpha = 1$	With benefit of screening		Without benefit of screening	
Strategy[1]	Mean	S.E.	Mean	S.E.
A4069	18.79	2.01	11.55	1.01
B4068	14.24	0.95	11.40	0.70
A5069	17.55	2.03	10.86	1.00
B5068	13.41	1.05	10.86	0.81
$\alpha = 1.25$	With benefit of screening		Without benefit of screening	
Strategy[1]	Mean	S.E.	Mean	S.E.
A4069	13.52	1.16	8.70	0.66
B4068	11.13	0.63	8.99	0.52
A5069	12.99	1.07	8.26	0.67
B5068	10.62	0.71	8.60	0.60
$\alpha = 1.5$	With benefit of screening		Without benefit of screening	
Strategy[1]	Mean	S.E.	Mean	S.E.
A4069	10.92	0.74	7.07	0.51
B4068	9.40	0.49	7.57	0.42
A5069	10.65	0.75	6.74	0.51
B5068	8.98	0.55	7.24	0.47

[1] A4069: Annual exams in the age interval 40-69 years. B4068: Biennial exams in the age interval 40-68 years. A5069: Annual exams in the age interval 50-69 years. B5068: Biennial exams in the age interval 50-68 years.

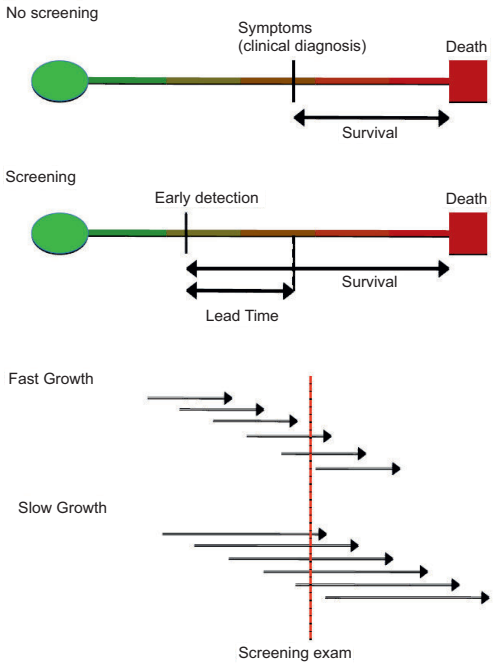


Figure A.1: *Lead-time (top) and length bias (bottom).*

Table A.5: *Mortality summary. One hundred simulated scenarios for each screening strategy. Time horizon 0-85 years.*

With benefit of screening								
$\alpha = 1$	Cumulative mortality		Deaths by BC		Deaths by BC SD[2]		Deaths by BC I[2]	
Strategy[1]	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
A4069	3.04	0.05	38.90	0.59	40.82	0.82	51.03	1.74
B4068	3.24	0.05	41.53	0.57	43.12	0.93	55.89	1.25
A5069	3.10	0.05	39.71	0.61	40.02	0.87	51.40	1.96
B5068	3.29	0.05	42.10	0.58	42.60	0.97	56.63	1.66
$\alpha = 1.25$	Cumulative mortality		Deaths by BC		Deaths by BC SD[2]		Deaths by BC I[2]	
Strategy[1]	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
A4069	3.13	0.05	40.10	0.60	40.06	0.88	57.47	1.80
B4068	3.34	0.05	42.73	0.60	41.36	0.95	61.32	1.31
A5069	3.19	0.05	40.92	0.61	39.17	0.91	57.85	1.99
B5068	3.38	0.05	43.30	0.60	40.84	0.98	62.19	1.62
$\alpha = 1.5$	Cumulative mortality		Deaths by BC		Deaths by BC SD[2]		Deaths by BC I[2]	
Strategy[1]	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
A4069	3.19	0.05	40.88	0.61	39.24	0.88	62.60	1.72
B4068	3.40	0.05	43.54	0.62	39.79	0.97	65.73	1.22
A5069	3.25	0.05	41.69	0.61	38.26	0.88	63.00	1.90
B5068	3.44	0.05	44.10	0.62	39.24	0.97	66.66	1.59
Without benefit of screening								
$\alpha = 1$	Cumulative mortality		Deaths by BC		Deaths by BC SD[2]		Deaths by BC I[2]	
Strategy[1]	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
A4069	3.72	0.05	47.65	0.59	57.88	0.81	59.78	1.73
B4068	3.72	0.05	47.65	0.59	58.10	0.90	59.97	1.23
A5069	3.72	0.05	47.65	0.59	57.93	0.86	60.78	1.91
B5068	3.72	0.05	47.65	0.59	58.23	0.97	60.88	1.53
$\alpha = 1.25$	Cumulative mortality		Deaths by BC		Deaths by BC SD[2]		Deaths by BC I[2]	
Strategy[1]	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
A4069	3.80	0.05	48.72	0.57	56.78	0.80	66.40	1.73
B4068	3.80	0.05	48.72	0.57	55.96	0.88	65.50	1.28
A5069	3.80	0.05	48.72	0.57	56.74	0.83	67.34	1.79
B5068	3.80	0.05	48.72	0.57	56.07	0.96	66.46	1.60
$\alpha = 1.5$	Cumulative mortality		Deaths by BC		Deaths by BC SD[2]		Deaths by BC I[2]	
Strategy[1]	Mean	S.E.	Mean	S.E.	Mean	S.E.	Mean	S.E.
A4069	3.86	0.05	49.49	0.58	55.89	0.77	71.71	1.67
B4068	3.86	0.05	49.49	0.58	54.22	0.85	69.96	1.17
A5069	3.86	0.05	49.49	0.58	55.76	0.83	72.72	1.65
B5068	3.86	0.05	49.49	0.58	54.27	0.90	71.07	1.45

[1] A4069: Annual exams in the age interval 40-69 years. B4068: Biennial exams in the age interval 40-68 years. A5069: Annual exams in the age interval 50-69 years. B5068: Biennial exams in the age interval 50-68 years.

[2] SD: Screen-detected cases, I: Interval cases.

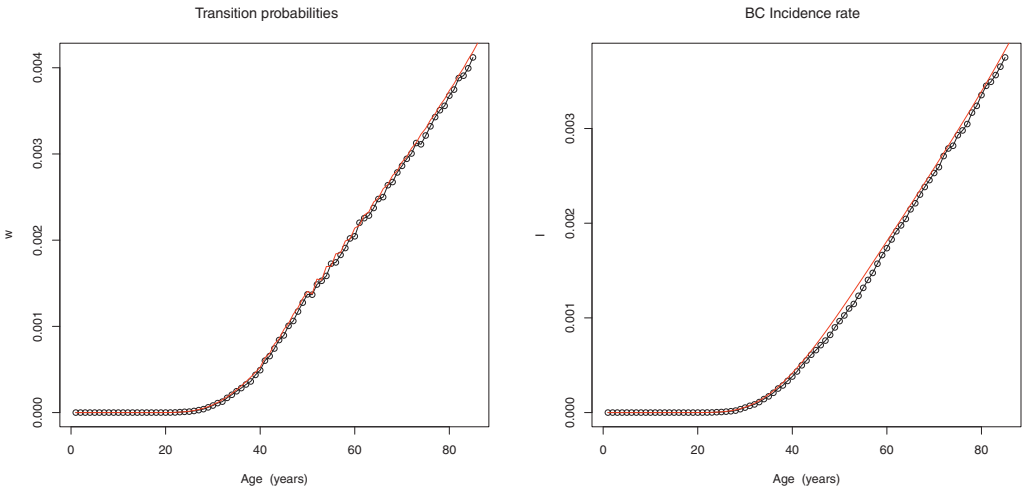


Figure A.2: *Mean simulated (black dots) and observed (red line) BC transition to Sp and incidence rates. One hundred simulated scenarios for each screening strategy. Time horizon 0-85 years.*

References

Blanch, J., Sala, M., Ibáñez, J., Domingo, L., Fernandez, B., Otegi, A., Barata, T., Zubizarreta, R., Ferrer, J., Castells, X., Rué, M. and Salas, D. (2014). Impact of risk factors on different interval cancer subtypes in a population-based breast cancer screening programme. *PLoS One*, 9(10): e110207.

Borras, J., Ameijide, A., Vilardell, L., Valls, J., Marcos-Gragera, R. and Izquierdo, A. (2008). Trends in cancer incidence in Catalonia, 1985-2002. *Medicina Clínica (Barcelona)*, 131 Suppl 1, 11–18.

Burton, A., Altman, D. G., Royston, P. and Holder, L. (2006). The design of simulation studies in medical statistics. *Statistics in Medicine*, 25, 4279–4292.

Bush, D., Smith, B., Younger, J. and Michaelson, J. S. (2010). The non-breast-cancer death rate among breast cancer patients. *Breast Cancer Research and Treatment*.

Domingo, L., Salas, D., Zubizarreta, R., Baré, M., Sarriugarte, G., Barata, T., Ibáñez, J., Blanch, J., Puig-Vives, M., Fernández, A., Castells, X. and Sala, M. (2014). Tumor phenotype and breast density in distinct categories of interval cancer: results of population-based mammography screening in Spain. *Breast Cancer Research*, 16: R3.

Duffy, S. W., Nagtegaal, I. D., Wallis, M., Cafferty, F. H., Houssami, N., Warwick, J., Allgood, P. C., Kearins, O., Tappenden, N., O'Sullivan, E. and Lawrence, G. (2008). Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. *American Journal of Epidemiology*, 168, 98–104.

Galceran, J., Puigdefabregas, A., Ribas, G., Izquierdo, A., Pareja, L. and Marcos-Gragera, R. (2008). Cancer survival trends in Catalonia and comparison with Europe. *Medicina Clínica (Barcelona)*, 131 Suppl 1, 19–24.

Gotzsche, P. C. and Nielsen, M. (2009). Screening for breast cancer with mammography. *Cochrane Database of Systematic Reviews*, (4), CD001877.

Kafadar, K. and Prorok, P. C. (2009). Effect of length biased sampling of unobserved sojourn times on the survival distribution when disease is screen detected. *Statistics in Medicine*, 28, 2116–2146.

Lee, S. and Zelen, M. (2006). A stochastic model for predicting the mortality of breast cancer. *Journal of the National Cancer Institute Monographs*, 79–86.

Lee, S. J. and Zelen, M. (1998). Scheduling periodic examinations for the early detection of disease: applications to breast cancer. *Journal of the American Statistical Association*, 93, 1271–1281.

Lee, S. J. and Zelen, M. (2008). Mortality modeling of early detection programs. *Biometrics*, 64, 386–395.

Lehtimaki, T., Lundin, M., Linder, N., Sihto, H., Holli, K., Turpeenniemi-Hujanen, T., Kataja, V., Isola, J., Joensuu, H. and Lundin, J. (2011). Long-term prognosis of breast cancer detected by mammography screening or other methods. *Breast Cancer Research and Treatment*, 13, R134.

Mahnken, J. D., Chan, W., Jr Freeman, D. H. and Freeman, J. L. (2008). Reducing the effects of lead-time bias, length bias and over-detection in evaluating screening mammography: a censored bivariate data approach. *Statistical Methods in Medical Research*, 17, 643–663.

Martinez-Alonso, M., Vilaprinyo, E., Marcos-Gragera, R. and Rue, M. (2010). Breast cancer incidence and overdiagnosis in Catalonia (Spain). *Breast Cancer Research and Treatment*, 12, R58.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Shen, Y., Yang, Y., Inoue, L. Y., Munsell, M. F., Miller, A. B. and Berry, D. A. (2005). Role of detection method in predicting breast cancer survival: analysis of randomized screening trials. *Journal of the National Cancer Institute*, 97, 1195–1203.

Trivedi, P. K. and Zimmer, D. M. (2007). Copula modeling: an introduction for practitioners. *Foundation and trends in Econometrics*, 1, 1–111.

Vilaprinyo, E., Gispert, R., Martinez-Alonso, M., Carles, M., Pla, R., Espinas, J. A. and Rue, M. (2008). Competing risks to breast cancer mortality in Catalonia. *BMC Cancer*, 8, 331.

Vilaprinyo, E., Rue, M., Marcos-Gragera, R. and Martinez-Alonso, M. (2009). Estimation of age- and stage-specific Catalan breast cancer survival functions using US and Catalan survival data. *BMC Cancer*, 9, 98.

Walter, S. D. and Day, N. E. (1983). Estimation of the duration of a pre-clinical disease state using screening data. *American Journal of Epidemiology*, 118, 865–886.

Walter, S. D. and Stitt, L. W. (1987). Evaluating the survival of cancer cases detected by screening. *Statistics in Medicine*, 6, 885–900.

Xu, J. L., Fagerstrom, R. M. and Prorok, P. C. (1999). Estimation of post-lead-time survival under dependence between lead-time and post-lead-time survival. *Statistics in Medicine*, 18, 155–162.

Xu, J. L. and Prorok, P. C. (1995). Non-parametric estimation of the post-lead-time survival distribution of screen-detected cancer cases. *Statistics in Medicine*, 14, 2715–2725.

Zelen, M. and Feinleib, M. (1969). On the theory of screening for chronic diseases. *Biometrika*, 56, 601–614.