Assessment of lead-time bias in estimates of relative survival for breast cancer

Therese M-L Andersson[a,*], Mark J Rutherford[b], Keith Humphreys[a]

[a]*Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden*
[b]*Department of Health Sciences, University of Leicester, UK*

Abstract

Relative survival ratios (RSRs) can be useful for evaluating the impact of changes in cancer care on the prognosis of cancer patients or for comparing the prognosis for different subgroups of patients, but their use is problematic for cancer sites where screening has been introduced due to the potential of lead-time bias. Lead-time is survival time that is added to a patient's survival time because of an earlier diagnosis irrespective of a possibly postponed time of death. In the presence of screening it is difficult to disentangle how much of an observed improvement in survival is real and how much is due to lead-time bias. Even so, RSRs are often presented for breast cancer, a site where screening has led to early diagnosis, with the assumption that the lead-time bias is small. We describe a simulation-based framework for studying the lead-time bias due to mammography screening on RSRs of breast cancer based on a natural history model developed in a Swedish setting. We have performed simulations, using this framework, under different assumptions for screening sensitivity and breast cancer survival with the aim of estimating the lead-time bias. Screening every second year among ages 40-75 was introduced assuming that screening had no effect on survival, except for lead-time bias. Relative survival was estimated both with and without screening to enable quantification of the lead-time bias. Scenarios with low, moderate and high breast cancer survival, and low, moderate and high screening sensitivity were simulated, and the lead-time bias assessed in all scenarios.

Keywords: Lead time, Relative survival, Breast cancer, Mammography screening, Simulation study

1. Introduction

Cancer patient survival is often used as a measure of cancer outcome [1, 2], and is frequently reported in the scientific literature [3, 4] as well as in policy documents and cancer registry reports [5, 6]. 5-year relative survival ratios (RSRs) are especially often reported, but also 1-year and 10-year RSRs. The 1-, 5- and 10-year RSRs provide,

---

*Corresponding author: Therese M-L Andersson, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Box 281, SE-17177 Stockholm, Sweden. E-mail: therese.m-l.andersson@ki.se

under certain assumptions, estimates of the proportion of cancer patients that would survive 1, 5 and 10 years, respectively, if the studied cancer was the only possible cause of death [7]. Relative survival is estimated by contrasting the all-cause survival among the cancer patients to the survival in a comparable group in the general population [1]. By comparing RSRs over calendar time or between groups one can draw conclusions about changes in the prognosis of cancer patients or differences in prognosis between groups. Even though temporal trends in RSRs can be useful for evaluating the impact of changes in cancer care on the prognosis of cancer patients, their use has been criticised, mainly because of the potential for lead-time bias [8, 9]. As far as we are aware, the impact of lead time bias on population-based measures of cancer patient survival, such as RSRs, has not previously been studied, although quantification of lead time bias due to mammography screening has been of interest in many studies that aim to assess the effectiveness of screening [10, 11, 12].

Lead-time is survival time that is added to a patient's survival time because of an earlier diagnosis irrespective of a possibly postponed time of death [13, 14]. In the presence of lead-time, the survival times of the patients are prolonged by an early diagnosis, and the survival proportion at any given time point is therefore increased even if no real improvement in survival is experienced. On the other hand, early diagnosis of cancer can lead to a real improvement of survival. In the presence of screening it is difficult to disentangle how much of an observed improvement in survival is real and how much is due to lead-time bias. Even so, RSRs are often presented for breast cancer [15], a site where screening has led to early diagnosis, with the assumption that the lead-time bias is small. Tryggvadottir *et. al.* [15] show that, in all Nordic countries, between 1964-2003, both the 5- and 10-year RSR increased by 20-30 percentage points. Many changes in cancer treatment and care were introduced during this period, and the reason for the observed increase is likely to be multifactorial. It is difficult to know how much of the observed improvement in survival over these years is due to lead time bias, especially because mammography screening was introduced gradually during this period and because sensitivity has increased gradually over time. For Denmark the 5-year RSR was 5-10 percentage points lower than the other Nordic countries. Tryggvadottir *et. al.* suggested that this could in part be explained by the late introduction of national organised screening in Denmark, but did not discuss the possibility that the difference could also be due to lead-time bias.

The aim of the present study is to understand more about lead-time bias due to mammography screening in estimates of 1-, 5- and 10-year age-standardised RSRs of breast cancer, using a novel simulation approach based on random effects tumour growth modelling. We investigate nine different scenarios with different assumptions about screening sensitivity and breast cancer survival. The paper is laid out as follows. In Section 2 we describe relative survival methodology. In Section 3 we describe the statistical models of natural history used for the simulation. Section 4 describes the simulation and statistical analysis in detail, and results are presented in Section 5. The paper ends with a discussion in Section 6.

2. Relative survival

The method of choice for estimating cancer patient survival in a population-based setting is relative survival [1, 7], which is defined as the observed (all-cause) survival among the cancer patients divided by the expected survival the patients would have experienced had they not had cancer. The expected survival is typically obtained from nationwide population mortality rates, stratified by age, sex and calendar year. Relative survival aims at estimating the net survival, interpreted as the proportion of patients still alive at a certain point after diagnosis, in the hypothetical scenario where the cancer of interest is the only possible cause of death. For most types of cancer the relative survival has increased over the last few decades, indicating that cancer treatment has improved. The main reason for using a relative survival approach is that it does not rely on correct classification of cause of death. The cause of death can be poorly reported, especially among elderly patients [16], and even when the reporting is good it can be difficult to determine if the death of a cancer patient is due to the cancer of interest or not (for example death from treatment complications) [17].

In the relative survival setting, the overall survival, $S(t)$, as a function of time t since diagnosis, can be written as

$$S(t; z) = S^*(t; z')R(t; z), \tag{1}$$

where $R(t)$ represents the relative survival and $S^*(t)$ is the survival the patient would have been expected to experience had they not had cancer. The overall as well as the relative survival can vary by values of the covariates, z, which can, for example, be patient characteristics such as sex, age and calendar year of diagnosis and/or tumour characteristics such as stage or grade. The expected survival is allowed to vary by the factors z', on which the population mortality rates are stratified, which usually represent a subset of z.

The hazard analogue of relative survival is excess hazard, and it measures the mortality the patients experience in excess of what would have been expected if they had not had cancer. The overall hazard, $h(t)$, among the patients is written as the sum of the expected hazard, $h^*(t)$, and the excess hazard, $\lambda(t)$, associated with the cancer

$$h(t; z) = h^*(t; z') + \lambda(t; z). \tag{2}$$

Traditionally, RSRs have been estimated non-parametrically using a life-table approach [18, 19, 20], and the excess hazard has often been modelled using Poisson regression using a link function that takes the expected mortality into account [21]. However, it has recently been suggested that RSRs should be estimated using a modelling approach that enables flexible modelling of the baseline excess hazard [7], and one such model is the flexible parametric survival model [22, 23].

2.1. Flexible parametric survival model

The flexible parametric survival model [22, 23] uses restricted cubic splines to model the baseline cumulative hazard. The use of splines enables the model to capture complex baseline cumulative hazard functions, and gives a parametric model without the need of strong distributional assumptions. The flexible parametric survival model

3

was first introduced by Royston and Parmar in 2001 [24, 22]. The model has also been extended for relative survival, by modelling the log cumulative excess hazard using restricted cubic splines [25, 23]. This extension is described below.

By integrating equation (2), the overall cumulative hazard, $H(t)$, can be expressed as

$$H(t; z) = H^*(t; z^{'}) + \Lambda(t; z), \tag{3}$$

where $H^*(t)$ is the cumulative expected hazard and $\Lambda(t)$ is the cumulative excess hazard. In a flexible parametric survival model for relative survival, the log cumulative excess hazard, $\ln \Lambda(t; z)$, is modelled as a function of follow-up time, t, using splines as:

$$\ln(\Lambda(t)) = s(x; \boldsymbol{\gamma_0}) \tag{4}$$

where $x = \ln(t)$ and $s(x; \boldsymbol{\gamma_0})$ is a restricted cubic spline function. The latter is defined as

$$s(x; \boldsymbol{\gamma_0}) = \gamma_{00} + \gamma_{01} v_1(x) + \gamma_{02} v_2(x) + ... + \gamma_{0K-1} v_{K-1}(x), \tag{5}$$

where K is the number of knots and the pth basis function is defined as

$$v_p(x) = \begin{cases} x, & \text{for } p = 1 \\ (x - k_p)_+^3 - \phi_p(x - k_1)_+^3 - (1 - \phi_p)(x - k_K)_+^3, & \text{for } p = 2, ..., K - 1 \end{cases} \tag{6}$$

where $u_+ = u$ if $u > 0$ and $u_+ = 0$ if $u \leq 0$, k_1 is the position of the first knot, k_K the position of the last knot, k_p the position of the pth knot, and $\phi_p = \frac{k_K - k_p}{k_K - k_1}$. The AIC and BIC can be used as guidance when deciding on the number of knots used and the placement of the knots. The model has been shown to be robust to the number and placement of knots [26].

Covariates, z, can be introduced when modelling the log cumulative excess hazard;

$$\ln(\Lambda(t; z)) = s(x; \boldsymbol{\gamma_0}) + z\boldsymbol{\beta}, \tag{7}$$

and the relative survival can easily be obtained using the relationship between the survival and the cumulative hazard function,

$$R(t; z) = \exp(-\Lambda(t; z)) = \exp(-\exp(s(x; \boldsymbol{\gamma_0}) + z\boldsymbol{\beta})). \tag{8}$$

2.2. Age-standardised relative survival ratio

One of the most commonly reported measures of population-based cancer patient survival is the age-standardised 5-year relative survival ratio (RSR). It can be obtained as a weighted average of the predicted relative survival from a flexible parametric model,

$$R_s(5) = \frac{1}{n} \sum_{i=1}^{n} \hat{R}_i(5) * w_i, \tag{9}$$

where n is the number of observations, $\hat{R}_i(5)$ is the predicted 5-year RSR for individual i and w_i is a weight for individual i. The weights are calculated using age distributions in standard reference populations as $w_i = w_i^s / a_i$, where a_i is the proportion in the age group to which the individual i belongs and w_i^s is the proportion belonging to that age group in the reference population. The age-standardised 1-year and 10-year RSRs can be estimated similarly.

4

3. The natural history model used for the simulation

To quantify the contribution of lead time to RSRs, we developed a simulation approach. As basis we generated birth cohorts of women and then added onset of disease based on the probabilities of developing breast cancer at different ages. In all simulations, for simplicity, we allowed only one breast cancer per individual. Independently of the onset of breast cancer, time to death due to other causes than breast cancer was simulated from birth based on mortality rates in the Swedish population. For individuals with an onset of breast cancer (before time of death), prior to assigning a diagnosis date we simulated the growth of the tumour. Tumours were assumed to grow exponentially in size until detection, with growth rates differing across individuals. Individuals were subsequently monitored under two scenarios, without and with mammography screening. In the absence of screening, age and tumour size at diagnosis were recorded for each individual based on formulating a model for time (from cancer onset) to symptomatic detection. Time to death due to breast cancer was measured from age at symptomatic diagnosis and assumed to follow a Weibull distribution, with age at (symptomatic) diagnosis as a covariate.

We subsequently imposed a screening programme on the cohorts where screening sensitivity depended on the size of the tumour presented at screening. With a screening programme imposed, some tumours are diagnosed early (are screen detected) while others are diagnosed symptomatically. For all tumours, we calculated the age and size of tumour at diagnosis, with and without screening. All steps of the simulation are described in more detail below.

We compared relative survival estimates calculated from the simulated data without imposed screening, to estimates based on the same sets of individuals, but with screening imposed.

3.1. Onset of breast cancer

Annual birth cohorts from 1870-1965 were simulated with 10 000 women in each cohort, which is approximately a fifth of the observed size of birth cohorts of females in Sweden. We used incidence rates of breast cancer in Sweden in 5-year age categories from 1973 to derive probabilities of cancer onset which we could use in our simulation. The year 1973 was chosen because this was prior to the introduction of mammography screening in Sweden [27]. By assuming that all tumours were initiated 10 years prior to diagnosis, we converted incidence rates to probabilities of cancer onset. This assumption was made only to derive cancer onset probabilities. In all simulations, times between cancer onset and diagnosis varied between individuals; see Section 3.2. The highest age category for incidence was 85+, and the converted probability for this group was used up until age 99. The lowest age category with an observed incidence above 0 was 20-25 years, so the onset of disease could therefore occur between the ages of 10 and 99.

3.2. Tumour growth and clinical detection

For individuals simulated with an onset of breast cancer, age at symptomatic diagnosis (in the absence of a screening program) and size at diagnosis were simulated using the tumour growth model and procedure described below.

We first generated an inverse growth rate (1/growth rate), r, for each individual from a Gamma distribution with shape parameter τ_1, inverse scale parameter τ_2 and density function

$$\frac{\tau_2^{\tau_1}}{\Gamma(\tau_1)} r^{\tau_1-1} \exp(-\tau_2 r), r \geq 0. \tag{10}$$

Each tumour was assumed to be spherical and to grow exponentially in volume, so that a tumour with inverse growth rate r would have a volume, t years after onset, of

$$V(t) = V_{cell} \exp(t/r), \tag{11}$$

where V_{cell} is the volume of one cell (i.e. the volume of the tumour at onset). We chose V_{cell} to be the volume of the tumour when its diameter was 0.5 mm, representing a tumour size which is unlikely to be detected at screening.

We then allowed time from cancer onset to symptomatic detection, T_{det}, to depend on tumour volume according to the hazard function

$$P(T_{det} \in [t, t+dt]|T_{det} > t) = \eta V(t)dt + o(dt), \tag{12}$$

where η is a constant.

With the above choices of models for tumour growth and symptomatic detection, it can be shown that the conditional distribution of tumour volume at symptomatic detection, V_{det}, given a growth rate, can be written explicitly as

$$F_{V_{det}|R}(v) = 1 - \exp(-\eta r(v - V_{cell})), v \geq V_{cell}, \tag{13}$$

and that time from onset to symptomatic detection, T_{det}, which is a fixed function of r, V_{det} and V_{cell}, can be calculated by,

$$T_{det} = r \ln(V_{det}/V_{cell}). \tag{14}$$

This tumour growth model has been described and applied by Plevritis *et. al.* [28] and by Abrahamsson and Humphreys [29]. To choose values of the parameters, τ_1, τ_2 and η, we used data on the tumour size distribution for breast cancers diagnosed during the years 1977-1979 from the Stockholm-Gotland regional quality register. Maximum likelihood estimation was used, based on the unconditional distribution of tumour sizes (in the absence of screening) under the above model (Section 3.1 of Plevritis *et. al.* [28]). For model identifiability we assumed $\tau_1 = \tau_2$, as in Plevritis *et. al.* [28]. The maximum likelihood estimates of the parameters were $\tau_1 = \tau_2 = 1.385$ and $\eta = 0.0002566$, and these values were used in our simulations.

3.3. Death due to breast cancer and other causes

Independently of the onset of breast cancer, time to death due to other causes than breast cancer was simulated from birth, based on mortality rates in the Swedish population by 1-year age and calendar year categories, and assuming that the longest possible lifetime is 100 years. Mortality rates for years 1870-2011 were obtained from the Human Mortality database [30] and mortality rates beyond year 2011 were obtained from mortality projections created by Statistics Sweden [31].

Our next step was to simulate death due to breast cancer. Survival times were assumed to follow a Weibull distribution, with age at symptomatic diagnosis as a covariate with linear effect. We therefore used the hazard function

$$h(t) = \gamma \alpha t^{\alpha-1} \exp(\rho * \text{age}). \tag{15}$$

Values for the shape, α, and scale, γ, parameters for the Weibull distribution, as well as the log hazard ratio for age, ρ, were found by fitting a Weibull model to data from the Swedish Cancer Registry on breast cancers diagnosed in Sweden from 1970-1974. Even though there are a range of clinically important factors for breast cancer survival, such as treatment and stage of disease, we did not consider these in our model. The reason for this was that these factors are rarely considered in population-based measures of cancer patient survival when the purpose is to monitor trends, or to estimate the overall age-standardised relative survival. The age at death from breast cancer, for each individual with cancer onset, was calculated by adding their simulated survival time from breast cancer to their age at symptomatic detection. The final age at death for each individual was finally defined as the minimum of age at death due to breast cancer and death due to other causes.

We simulated three different scenarios for time to death due to breast cancer using the estimated scale and shape parameters and their upper and lower confidence limits, providing scenarios with low, moderate and high breast cancer survival. A log hazard ratio, ρ, of 0.021 for age (obtained by fitting the model to real data as described above) was assumed in all scenarios. The three scenarios correspond to a 5-year cause-specific survival, for a woman of age 60 at diagnosis, of 69%, 72% and 76% for low, moderate and high breast cancer survival, respectively.

3.4. Screening sensitivity

All women were assumed to participate in screening every second year from the age of 40 until the age of 74. At each screening round any existing tumours not yet diagnosed, were screen detected with probability calculated as

$$\frac{\exp(\beta_1 + \beta_2 d)}{1 + \exp(\beta_1 + \beta_2 d)}. \tag{16}$$

where d represents tumour diameter. This logistic function has also been used by Abrahamsson and Humphreys [29] who estimated parameter values (for screening sensitivity) based on fitting a continuous growth model, as described in equations (10)-(12), to data on tumour size and screening history of Swedish women with postmenopausal breast cancer [32]. Abrahamsson and Humphreys [29] allowed screening sensitivity to depend on mammographic density as well as tumour size (i.e. used $\beta_1 + \beta_2 d + \beta_3 m$ as a linear predictor with a logistic function, where m is mammographic density). We used their point estimate for β_2 for our β_2 and for β_1 we used $\hat{\beta_1} + \hat{\beta_3} 0.15$ from their paper, so that the intercept for screening sensitivity represents a woman with 15 percent mammographic density. This was used as a scenario of moderate screening sensitivity in our simulations. Similarly, we used the upper and lower confidence limits of the parameters $\hat{\beta_1}$ and $\hat{\beta_2}$ in Abrahamsson and Humphreys [29] to simulate scenarios of low and high screening sensitivity. For a tumour of 12 mm, for example, the values we

used in our simulations correspond to screening sensitivities of 0.58, 0.84 and 0.96 for low, moderate and high sensitivities, respectively.

4. Methods

Combining three scenarios for time to death due to breast cancer with three scenarios for screening, we explored nine different settings for our simulations, which are summarized in Table 1. The simulations were performed by first simulating the data without screening for each of the three scenarios for time to death, and then applying the three different screening sensitivity scenarios to the three scenarios for time to death. For each setting we simulated 200 datasets. For each dataset we estimated the 1-, 5- and 10-year age-standardised RSR for women diagnosed during 1970-1974, both in the presence and in the absence of screening. Since the actual survival time is not changed by screening/early detection in our simulation, the only difference in the two estimates of relative survival is due to the earlier diagnosis for those women diagnosed at screening. We could therefore estimate the lead-time bias as well as the average lead-time among those that are screen detected. We calculated both absolute and relative lead-time bias.

To estimate the 1, 5 and 10-year age standardised RSRs we used a flexible parametric survival model with 5 degrees of freedom (d.f.) for the baseline log cumulative excess hazard and included age in the model as a linear effect, allowing for non-proportional hazards by introducing a 3 d.f. interaction with time. This is an overparameterised model since the simulation is based on a Weibull model, but this will not affect the estimates of lead-time bias. Survival estimates from flexible parametric models have been shown to be robust to the number of knots [26], and we have therefore chosen to use the amount of flexibility that would normally be used for these types of data. The cancer patients were followed from diagnosis to date of death or censored 12 years after diagnosis, and the analysed time-scale was time since diagnosis. In this simulation setting we have full follow-up, but have chosen to censor individuals after 12 years to mimic the situation of administrative censoring that is often the case with these type of data. Also, since the excess mortality decreases over time since diagnosis the model is less robust and there could potentially be problem with convergence if including the full time frame where the excess mortality is low. Age-standardisation was carried out using international cancer survival standards (ICSS) weights [33] where the age groups min-44, 45-54, 55-64, 65-74 and 75 and above correspond to 7%, 12%, 23%, 29% and 29% of the standard reference population, respectively.

4.1. Data validation

For our main validation of the simulation strategy we compared the simulated data from setting 5 (moderate breast cancer survival and moderate screening sensitivity) to data from the Swedish Cancer Registry and the Stockholm-Gotland regional quality register for breast cancer. As in the main analysis, we restricted the validation to those diagnosed between 1970-1974 and made comparisons to the same years in the national register data. For the simulated data we present the averages from the 200 simulations. Due to lack of information on tumour size at diagnosis in the Swedish cancer register

8

during this period we used data from Stockholm-Gotland regional quality register for breast cancer for the years 1977-1979 for this comparison. The simulated data gave a slightly older age distribution than that seen in the register data (Table 2), but the proportion of (all cause) deaths within 12 years corresponds well with the observed data and the number of diagnoses in the register was approximately 5 times the number of diagnoses in the simulated data, which is as expected, since the birth cohorts of women in Sweden are approximately 5 times larger than the simulated birth cohorts. The distribution of tumour size at diagnosis corresponds well with the registry data, although the tumours in the simulated data are slightly larger. For the scenario with screening we compared the simulated data to data from 1996-1998 in the Swedish Cancer Registry and the Stockholm-Gotland regional quality register for breast cancer. We saw a clear shift towards smaller tumours from the period 1977-1979 to the period 1996-1998 compared to the 1977-1979 data,which was also observed in our simulated data. Comparisons between the registry data and all simulation settings (1-9) are presented in supplementary online tables.

5. Results

The proportion of cancer cases that were detected at screening ranged from 30% to 46% in the low and high screening sensitivity scenarios, respectively (Table 3). However the mean and median lead-time among screen detected cancers did not vary greatly between the different scenarios (Table 3). With moderate screening the mean lead-time among screen detected was 2.29 years and the median lead-time was 1.31 years, showing that the majority of breast cancers that are detected at screening would be detected symptomatically within the coming 2 years. With a higher screening sensitivity the mean and median lead-times would instead be 2.76 and 1.67 years. All numbers in Table 3 are averages based on 200 simulations of each screening scenario. Breast cancer survival does not affect the amount of lead-time or the screening sensitivity, and therefore we only include settings 4-6 in the table since the results would be the same for the scenarios with low and high survival (except for differences due to random variation).

Table 4 shows the estimated 1, 5 and 10-year RSRs with and without screening as well as the absolute and relative differences between the estimates with and without screening, for the nine different simulation settings. All estimates are averages from 200 simulations of each setting. Both the absolute and relative bias increase with improved screening sensitivity and reduced cancer survival. The absolute and relative biases were generally larger for the 5-year estimates than the 1- and 10-year estimates. The largest absolute bias was 5.7 percentage units, which was observed for the 5-year RSR in the setting with low breast cancer survival and high screening sensitivity. The largest relative bias was 8.4 % for the 5-year RSR in the same setting.

6. Discussion

We have described a simulation-based approach for studying the lead-time bias due to mammography screening on estimates of 1, 5 and 10-year age-standardised RSRs of breast cancer, and have estimated this bias under different assumptions for screening

sensitivity and different levels of breast cancer survival. The largest absolute bias was 5.7 percentage points and the largest relative bias was 8.4%. It is noteworthy that the largest bias, both absolute and relative, was observed for the 5-year RSR, which is also the measure which is most often reported. It is likely that this is due to a combination of there being a high short term survival (most patients survive the first year after diagnosis) and that when estimating long term survival (i.e. 10-year relative survival) the relative impact of lead time is modest.

The results from our study suggest that lead-time bias introduced by mammography screening does not explain the survival improvement observed during the recent decades in the Nordic countries. The absolute as well as relative bias was generally small, and much smaller than the observed increase in relative survival between 1964-2003. However, in some settings the absolute bias reached 4.0-5.7 percentage points, on a survival around 68-77%, a difference that many would see as an interesting improvement in survival. This is comparable to the 5-10 percentage points lower 5-year RSR observed in Denmark compared to the other Nordic countries [15]. The late introduction of mammography screening was suggested as a potential explanation of this survival disadvantage among Danish women, but according to our results this could, at least to a sizeable extent, be due to lead-time bias.

The simulations used in this study were based on a previously published tumour growth model which has been shown to fit observed distributions of tumour size at diagnosis reasonably well [28, 29]. One of the main assumptions of the model is that tumours grow exponentially until detection. We compared the tumour size distribution in the simulated data to tumour sizes in register based data from the years 1977-1979 (prior to screening) and 1996-1998 (after the introduction of screening), and the simulated data seemed to fairly accurately describe the tumour sizes in the real data. The difference observed could, for example, be due to non-attendance in screening. We have assumed that all women regularly attend screening according to the screening program, but non-attendance could, in principle, be included in the simulations.

Our tumour growth model also assumes that all tumours will eventually show symptoms. There is, however, in reality a possibility that a proportion of diagnosed in situ tumours would never become invasive if they were not detected and left untreated. For this reason our analyses are all based on invasive tumours only. Our approach is also based on a specific model for screening sensitivity. This was also taken from published literature [29], and we implemented a screening program where women were invited every second year from the age of 40 to the age of 74. This is in line with the current screening program in Sweden, but the age limits has been different over time and in different counties within Sweden. Since we have assumed 100% adherence, which is of course not what is observed in practice, the true lead-time bias is likely to be lower than that observed in our simulations.

There are other biases associated with screening, and the most important in this setting is over-diagnosis. Over-diagnosis is often defined as the detection of tumours that, if left untreated, would not lead to symptoms or harm to the individual within the individual's lifetime [13, 14]. Over-diagnosis can occur if there are tumours that lack true malignant potential or if they are so slow-growing that the individual would die due to other causes before her life is threatened by the cancer. The latter will always occur to some extent when cancers are diagnosed earlier, and is seen in our

simulations. However, if there is a specific subtype of indolent cancer that is diagnosed due to screening, or if a great proportion of in situ cancers diagnosed due to screening lack malignant potential, the over-diagnosis can be substantial. In this study we only included invasive cancer, and have not considered over-diagnosis due to indolent or in situ cancer.

Lead-time can occur for other cancers sites than breast cancer and also without the introduction of a screening program, if cancers are diagnosed earlier in the natural history of the disease in some groups compared to others or across calendar time. Even though our results can be informative for the possible extent of lead-time bias on estimates of survival for other cancer sites, it is difficult to know how applicable they are for other sites. The bias will for example depend on the cancer patient survival, as seen in our simulations, which differs between cancers. However, the simulation approach used in this study can be used for other cancer sites if the tumour growth model and the model for screening sensitivity could be adapted to correspond to that specific cancer.

In summary, the possibility of lead-time bias should not be neglected when interpreting trends in breast cancer survival or differences between groups or countries. It is, however, not likely to have a great impact on survival estimates.

Acknowledgements

Funding

[1] Dickman PW and Adami HO. Interpreting trends in cancer patient survival. *J Intern Med* 2006; 260(2): 103–117.

[2] Maruvka YE, Tang M and Michor F. On the validity of using increases in 5-year survival rates to measure success in the fight against cancer. *PLoS One* 2014; 9(7): e83100.

[3] Verdecchia A, Francisci S, Brenner H et al. Recent cancer survival in Europe: a 2000-02 period analysis of EUROCARE-4 data. *Lancet Oncol* 2007; 8(9): 784–796.

[4] Björkholm M, Ohm L, Eloranta S et al. Success story of targeted therapy in chronic myeloid leukemia: a population-based study of patients diagnosed in Sweden from 1973 to 2008.

[5] Carstensen B, Storm HH and Schou G. *Survival of Danish Cancer Patients 1943-1987*, *APMIS*, volume 101 (Suppl. 33). Copenhagen: Danish Cancer Society, 1993.

[6] Berrino F, Capocaccia R, Estève J et al. *Survival of Cancer Patients in Europe: The EUROCARE-2 Study*. IARC Scientific Publications No. 151, Lyon: International Agency for Research on Cancer, 1999.

[7] Rutherford MJ, Dickman PW and Lambert PC. Comparison of methods for calculating relative survival in population-based studies. *Cancer Epidemiology* 2012; 36(1): 16–21.

[8] Measurement of progress against cancer. Extramural Committee to Assess Measures of Progress Against Cancer. *J Natl Cancer Inst* 1990; 82(10): 825–835.

[9] Welch HG, Schwartz LM and Woloshin S. Are increasing 5-year survival rates evidence of success against cancer? *JAMA* 2000; 283(22): 2975–2978.

[10] Walter SD and Day NE. Estimation of the duration of a pre-clinical disease state using screening data. *Am J Epidemiol* 1983; 118(6): 865–886.

[11] Duffy SW, Nagtegaal ID, Wallis M et al. Correcting for lead time and length bias in estimating the effect of screen detection on cancer survival. *Am J Epidemiol* 2008; 168(1): 98–104.

[12] Independent U K Panel on Breast Cancer Screening. The benefits and harms of breast cancer screening: an independent review. *Lancet* 2012; 380(9855): 1778–1786.

[13] Kramer BS and Croswell JM. Cancer screening: the clash of science and intuition. *Annu Rev Med* 2009; 60: 125–137.

[14] Croswell JM, Ransohoff DF and Kramer BS. Principles of cancer screening: lessons from history and study design issues. *Semin Oncol* 2010; 37(3): 202–215.

[15] Tryggvadottir L, Gislum M, Bray F et al. Trends in the survival of patients diagnosed with breast cancer in the nordic countries 1964-2003 followed up to the end of 2006. *Acta Oncologica* 2010; 49(5): 624–631.

[16] Fall K, Strmberg F, Rosell J et al. Reliability of death certificates in prostate cancer patients. *Scand J Urol Nephrol* 2008; 42(4): 352–357.

[17] Begg CB and Schrag D. Attribution of deaths following cancer treatment. *J Natl Cancer Inst* 2002; 94(14): 1044–1045.

[18] Ederer F and Heise H. Instructions to IBM 650 programmers in processing survival computations. methodological note no. 10, end results evaluation section. Technical report, National Cancer Institute, Bethesda, MD, 1959.

[19] Ederer F, Axtell L and Cutler S. The relative survival rate: A statistical methodology. *National Cancer Institute Monograph* 1961; 6: 101–121.

[20] Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 1982; 38(4): 933–942.

[21] Dickman PW, Sloggett A, Hills M et al. Regression models for relative survival. *Stat Med* 2004; 23(1): 51–64.

[22] Royston P and Parmar MKB. Flexible parametric proportional-hazards and proportional-odds models for censored survival data, with application to prognostic modelling and estimation of treatment effects. *Statistics in Medicine* 2002; 21(15): 2175–2197.

[23] Lambert PC and Royston P. Further development of flexible parametric models for survival analysis. *The Stata Journal* 2009; 9: 265–290.

[24] Royston P. Flexible parametric alternatives to the cox model, and more. *The Stata Journal* 2001; 1: 1–28.

[25] Nelson CP, Lambert PC, Squire IB et al. Flexible parametric models for relative survival, with application in coronary heart disease. *Stat Med* 2007; 26(30): 5486–5498.

[26] Rutherford MJ, Crowther MJ, Lambert PC. The use of restricted cubic splines to approximate complex hazard functions in the analysis of time-to-event data: a simulation study. *Journal of Statistical Computation and Simulation* 2015; 85:777–793.

[27] Jonsson H. *Evaluation of service screening with mammography in Sweden with special regard to its impact on breast cancer mortality*. PhD Thesis, Umeå Universitet, 2002.

[28] Plevritis SK, Salzman P, Sigal BM et al. A natural history model of stage progression applied to breast cancer. *Stat Med* 2007; 26(3): 581–595.

[29] Abrahamsson L and Humphreys K. A statistical model of breast cancer tumour growth with estimation of screening sensitivity as a function of mammographic density. *Stat Methods Med Res* 2016; 25: 1620-1637.

[30] The human mortality database. http://www.mortality.org/.

[31] Statistics Sweden (Statistiska Centralbyrån). The future population of Sweden 2012-2060. Technical report, Demographic reports 2012:2, 2012.

[32] Magnusson C, Baron J, Persson I et al. Body size in different periods of life and breast cancer risk in post-menopausal women. *Int J Cancer* 1998; 76(1): 29–34.

[33] Corazziari I, Quinn M and Capocaccia R. Standard cancer patient population for age standardising survival ratios. *Eur J Cancer* 2004; 40(15): 2307–2316.

Setting	Survival	Sensitivity	Scale (γ)	Shape (α)	β_1	β_2
1	Low	Low	0.032	0.75	-5.45	0.48
2	Low	Moderate	0.032	0.75	-5.04	0.56
3	Low	High	0.032	0.75	-4.67	0.65
4	Moderate	Low	0.028	0.74	-5.45	0.48
5	Moderate	Moderate	0.028	0.74	-5.04	0.56
6	Moderate	High	0.028	0.74	-4.67	0.65
7	High	Low	0.024	0.72	-5.45	0.48
8	High	Moderate	0.024	0.72	-5.04	0.56
9	High	High	0.024	0.72	-4.67	0.65

Table 1: Parameter values for breast cancer survival and screening sensitivity used in nine simulation settings.

	No screening		Screening	
	Simulated data	Register data	Simulated data	Register data
# diagnosed	3553	17228	3594	28715
Mean age at diagnosis	66	63	65	63
25th percentile of age	54	52	52	52
Median age	67	63	66	62
75th percentile of age	79	73	78	74
% dead within 12 years	65%	64%	63%	-
% size -17.5	39.5%	42.9%	59.3%	55.1%
% size 17.5-32.5	44.2%	44.9%	28.6%	35.3%
% size 32.5-47.5	11.8%	8.0%	8.2%	5.9%
% size 47.5-	4.5%	4.2%	3.9%	3.7%

Table 2: Comparison of simulated data (averages from 200 simulations) to data from the Swedish Cancer Registry; diagnoses 1970-1974 (no screening) & 1995-1999 (with screening). For size at diagnosis we compared to data from the Stockholm-Gotland regional quality register on breast cancer in the years 1977-1979 (no screening) and 1996-1998 (with screening).

Simulation setting	Sensitivity	Number of diagnoses with screening	% screen detected	Mean lead-time for screen detected (years)	Median lead-time for screen detected (years)
4	Low	3577 (3448-3688)	30.1 (28.8-31.7)	1.87 (1.69-2.05)	0.99 (0.90-1.10)
5	Moderate	3594 (3468-3714)	38.7 (37.3-40.5)	2.29 (2.13-2.45)	1.31 (1.21-1.43)
6	High	3612 (3487-3734)	45.7 (44.0-47.4)	2.76 (2.62-2.94)	1.67 (1.55-1.80)

Table 3: Proportion screen detected, and mean and median lead-time among screen detected cases in different simulation screening scenarios. All numbers are averages (with 2.5 and 97.5 percentiles in parenthesis) based on 200 simulations.

Simulation setting	1-year RSR no screening	1-year RSR screening	Absolute Bias	Relative Bias %
1	88.4 (87.0-89.6)	90.7 (89.5-91.9)	2.3 (1.9-2.8)	2.6 (2.1-3.2)
2	88.4 (87.0-89.6)	91.7 (90.5-92.8)	3.3 (2.8-3.9)	3.7 (3.1-4.5)
3	88.4 (87.0-89.6)	92.5 (91.1-93.7)	4.1 (3.5-4.8)	4.7 (4.2-5.4)
4	89.9 (88.9-90.9)	92.0 (91.0-93.0)	2.1 (1.6-2.5)	2.3 (1.8-2.7)
5	89.9 (88.9-90.9)	92.9 (91.8-93.8)	2.9 (2.3-3.4)	3.3 (2.6-3.8)
6	89.9 (88.9-90.9)	93.7 (92.6-94.6)	3.7 (3.1-4.4)	4.2 (3.4-4.9)
7	91.3 (90.4-92.3)	93.1 (92.2-94.1)	1.8 (1.4-2.2)	2.0 (1.5-2.5)
8	91.3 (90.4-92.3)	93.9 (93.0-94.9)	2.6 (1.9-3.1)	2.8 (2.1-3.4)
9	91.3 (90.4-92.3)	94.6 (93.5-95.6)	3.3 (2.7-3.9)	3.6 (2.9-4.3)

Simulation Setting	5-year RSR no screening	5-year RSR screening	Absolute Bias	Relative Bias %
1	68.1 (66.4-70.1)	70.6 (68.6-72.6)	2.5 (1.7-3.3)	3.7 (2.5-4.9)
2	68.1 (66.4-70.1)	72.1 (70.0-74.3)	4.0 (3.0-5.0)	5.8 (4.4-7.3)
3	68.1 (66.4-70.1)	73.8 (71.6-76.1)	5.7 (4.4-6.9)	8.4 (6.4-10.2)
4	72.5 (70.6-74.4)	74.7 (72.7-76.7)	2.2 (1.4-3.0)	3.0 (1.9-4.1)
5	72.5 (70.6-74.4)	76.1 (73.9-78.0)	3.5 (2.5-4.6)	4.9 (3.5-6.3)
6	72.5 (70.6-74.4)	77.6 (75.4-79.4)	5.1 (3.8-6.3)	7.0 (5.2-8.7)
7	77.1 (75.2-79.0)	79.0 (77.2-81.0)	1.9 (1.1-2.8)	2.5 (1.4-3.6)
8	77.1 (75.2-79.0)	80.1 (78.1-82.3)	3.1 (2.0-4.4)	4.0 (2.6-5.7)
9	77.1 (75.2-79.0)	81.5 (79.6-83.5)	4.5 (3.3-5.6)	5.8 (4.2-7.4)

Simulation setting	10-year RSR no screening	10-year RSR screening	Absolute Bias	Relative Bias %
1	53.8 (51.3-56.4)	55.4 (53.1-58.0)	1.5 (0.3-2.6)	2.9 (0.5-4.9)
2	53.8 (51.3-56.4)	56.3 (53.7-59.8)	2.6 (1.2-3.7)	4.6 (2.1-6.9)
3	53.8 (51.3-56.4)	57.6 (55.0-60.3)	3.8 (2.2-5.4)	7.1 (4.0-10.3)
4	60.0 (57.3-62.1)	61.4 (58.9-63.7)	1.4 (0.3-2.3)	2.3 (0.5-4.0)
5	60.0 (57.3-62.1)	62.3 (59.7-64.3)	2.3 (1.1-3.9)	3.8 (1.8-6.4)
6	60.0 (57.3-62.1)	63.4 (60.6-65.6)	3.4 (1.7-5.1)	5.7 (2.8-8.7)
7	66.8 (64.3-69.1)	68.0 (65.5-70.5)	1.2 (0.0-2.4)	1.8 (0.0-3.6)
8	66.8 (64.3-69.1)	68.8 (66.1-71.3)	2.0 (0.4-3.4)	3.0 (0.7-5.0)
9	66.8 (64.3-69.1)	69.8 (67.5-72.1)	3.0 (1.1-4.7)	4.5 (1.7-7.2)

Table 4: Estimated age-standardised RSRs and bias in the nine different simulated settings. All numbers are averages (with 2.5 and 97.5 percentiles in parenthesis) based on 200 simulations.