Cancer Epidemiology 46 (2017) 50–56

Contents lists available at ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

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

ARTICLE INFO

Article history:
Received 22 August 2016
Received in revised form 7 December 2016
Accepted 10 December 2016
Available online 24 December 2016

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

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.

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, 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 [10–12].

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

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

http://dx.doi.org/10.1016/j.canep.2016.12.004
1877-7821/

is small. Tryggvadottir et al. [15] show that, in all Nordic countries, between 1964 and 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

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

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, $\boldsymbol{z}$, 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 $\boldsymbol{z}'$, on which the population mortality rates are stratified, which usually represent a subset of $\boldsymbol{z}$.

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

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

Traditionally, RSRs have been estimated non-parametrically using a life-table approach [18–20], 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 was first introduced by Royston and Parmar in 2001 [22,24]. The model has also been extended for relative survival, by modelling the log cumulative excess hazard using restricted cubic splines [23,25]. This extension is described below.

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

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

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, $\ln\Lambda(t;\boldsymbol{z})$, 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

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

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

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

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 p^{th} knot, and $\phi_p = (k_K - k_p)/(k_K) - k_1$. 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, $\boldsymbol{z}$, can be introduced when modelling the log cumulative excess hazard;

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

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

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

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