EUROPEAN JOURNAL OF CANCER 47 (2011) 2626–2632

Available at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.ejconline.com

Estimating expected survival probabilities for relative survival analysis – Exploring the impact of including cancer patient mortality in the calculations

Mats Talbäck [a,b,*], *Paul W. Dickman* [b]

[a] *Unit for Epidemiology, Department of Statistics, Monitoring and Evaluation, National Board of Health and Welfare, SE-106 30 Stockholm, Sweden*

[b] *Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, P.O. Box 281, SE-171 77 Stockholm, Sweden*

ARTICLE INFO

Article history:
Available online 15 September 2011

Keywords:
Cancer survival
Relative survival
Survival probabilities
Excess mortality
Expected survival

ABSTRACT

Relative survival is a widely used measure of cancer patient survival, defined as the observed survival of the cancer patients divided by the expected survival of a comparable group from the general population, free from the cancer under study. In practise, expected survival is usually calculated from general population life tables. Such estimates are known to be biased since they also include mortality from the cancer patients, but the bias is ignored since mortality among individuals with a specific cancer is thought to constitute only a small proportion of total mortality. Using the computerised population registers that exist in Sweden we had the unique opportunity to calculate expected survival both including and excluding individuals with cancer, and thereby estimate the size of the bias arising from using general population estimates. We also evaluated a simple method to adjust expected survival probabilities estimated from general population statistics as an aid to researchers who do not have access to computerised registers of the entire national population.

Our results show that the bias is sufficiently small to be ignorable for most applications, notably for cancers with high or low mortality and for younger age groups (<60 years). However, the bias in relative survival estimates can be greater than 1 percent unit for older age groups for common cancers and even larger for all sites combined. For example, the bias in 10-year relative survival for men aged 75+ diagnosed with prostate cancer was 2.6 percent units, which we think is of sufficient magnitude to warrant adjustment.

1. Introduction

Relative survival is the preferred measure of cancer patient survival used by population-based cancer registries. It is, for example, used for comparing the survival among cancer patients in different countries or regions[1–3] as it enables comparisons of cancer patient survival while accounting for differences in general-population (non-cancer) mortality. The relative survival ratio is defined as the observed survival of the cancer patients divided by the expected survival of a comparable group from the general population, free from the cancer under study.[4] In their seminal 1961 paper, Ederer et al. [4] (p. 103) define 'the expected survival rate is that of a group similar to the patient group in such characteristics as

* *Corresponding author at:* Unit for Epidemiology, Department of Statistics, Monitoring and Evaluation, National Board of Health and Welfare, SE-106 30 Stockholm, Sweden. Tel.: +46 (0) 75 247 31 56, fax: +46 (0) 75 247 33 27.
E-mail address: mats.talback@socialstyrelsen.se (M. Talbäck).

0959-8049/$ - see front matter
doi:10.1016/j.ejca.2011.08.010

age, sex, and race, but *free of the specific disease under study*' [their italics].

However, the expected survival probabilities are usually calculated from general population life tables and since these tables reflect the mortality from all causes of death it is theoretically necessary to adjust the probabilities for cancer patient mortality. This is nonetheless rarely, if ever done, as the mortality from a specific form of cancer often is regarded as a negligible part of the total mortality and that general-population life table estimates provide a satisfactory proxy for a disease-free reference population. This seems plausible when the prevalence of the specific cancer is low (e.g. for younger patients and less common forms of cancer) but it could be questioned for older age groups, common forms of cancer, and for all sites combined.

Ederer et al. addressed this issue in 1961,[4] citing five studies performed during the 1950s, and concluded that 'since we are usually concerned with analysing survival of patients with specific forms of cancer, it appears that we do not need to make an adjustment in estimating expected survival from population life tables',[4] (p. 104). Oksanen addressed this issue in her 1998 doctoral thesis[5] but to our knowledge this issue has not otherwise been systematically evaluated. Oksanen, who modelled the relative survival and relative mortality for patients with prostate cancer, concluded that even though prostate cancer is the most common cancer in men the effect of excluding prevalent cases from the whole population in order to form a proper disease-free population had only a minor effect on the mortality figures.

Sweden maintains an electronic register of the entire population (currently around 9.4 million) along with registers of all incident cases of cancer and deaths. We therefore had the possibility to calculate expected survival (and relative survival) both including and excluding individuals with cancer from the population base, and thereby estimate the size of the bias arising from using general population estimates. We also evaluated a simple method to adjust expected survival probabilities estimated from general population statistics as an aid to researchers who do not have access to computerised registers of the entire national population.

2. Material and methods

2.1. Patients

This study was based on all cancer cases reported to the Swedish Cancer Registry between 1987 and 2001. Five common types of cancer and all sites combined were used in the analysis. A total of 286 thousand male and 279 thousand female cancers were included and patients were followed regarding censoring or death up to and including 31st December 2002. Complete follow-up, recorded as deceased or censored at the end of follow-up, was available for over 99% of the cases.

We selected five cancer types to represent different survival patterns among cancer patients. Colorectal cancer (ICD-10: C18-C21) was chosen to represent a common cancer with an intermediate patient survival, lung cancer (C34) represents a cancer with a short survival, skin cancer (C43-C44) a cancer with very favourable survival, breast cancer (C50) is the most common female cancer representing 25% of all female cancers at the beginning of the period and 29% at the end, prostate cancer (C61) is the most common male cancer representing 24% of all male cancers at the beginning of the period and 32% at the end. Prostate cancer also represents a cancer common in the older ages with a mean age at diagnosis of approximately 73 years during the study period. Survival for all sites combined (C00-C97) was also estimated since it is often reported.

Ninety-eight percent of the tumours recorded between 1987 and 2001 are histologically confirmed and an additional 1.6% verified by medical imaging. Patients diagnosed incidentally at autopsy or without any information regarding follow-up were excluded from the survival analysis. Only the first primary cancer at each site or groups thereof, as defined above, was included in the analyses and patients with multiple primary cancers diagnosed at different sites were included as independent entities.

Since 1958 every clinician, pathologist, and cytologist in Sweden is required by law to notify the Cancer Registry at the National Board of Health and Welfare of each new cancer diagnosed. The Cancer Registry is population based and from its inception the register has accumulated information on 2.3 million tumours for two million persons.

2.2. Statistical analysis

We used a database with all persons that have had a period of official residency in Sweden between 1986 and 2002 to determine the impact on the relative survival ratio (RSR) of including cancer patient mortality in the general population life tables. The database comprised information on 10.8 million persons regarding their residency at the end of each year, the date of death, and information regarding all diagnoses of cancer between 1958 and 2002. This database made it possible to censor, on an individual basis, each person after they received a specific diagnosis of cancer from the calculation of the expected survival probabilities. These expected probabilities are thus not affected by the mortality from patients diagnosed with the particular cancer under study, irrespective of whether the cancer in question is directly or indirectly related to the cause of death. Expected survival probabilities were also estimated from official population and mortality statistics[6] for the general population, which is the method most commonly used in relative survival analysis. We also evaluated a simple method to adjust these expected survival probabilities from the general population for cancer patient mortality as an aid to researchers who do not have access to computerised registers of the entire national population.

We estimated expected survival using the following three methods. Method 1 was regarded as the reference to which the other methods were compared.

1. Expected probabilities estimated from the database with individual population records where the cancer patients were censored at their time of diagnosis. This is the way expected survival should be calculated using the strict theoretical definition of relative survival.
2. Expected probabilities estimated from aggregated official population and mortality statistics without any correction for cancer patient mortality. This is the method commonly

used in practise even though it is known to be theoretically incorrect.

3. Estimates from Method 2 adjusted to account for the fact that deaths due to the cancer of interest are included. The adjusted expected survival probability is given by

$$P_{adj} = P_{gp}^{(1-\alpha)}$$

where P_{gp} is the expected survival probability estimated from general population mortality life tables (i.e. method 2) and α is the proportion of deaths due to cancer in the population. Further details of the method are provided in the appendix. We see that if $\alpha = 0$ then the standard method is unbiased and if $\alpha = 1$ (all deaths are due to cancer) then $P_{adj} = 1$ (everyone survives in the absence of cancer).

Adjustments were made separately for males and females. Application of the adjustment requires estimates of α, which should be possible to obtain from official cause-of-death statistics (i.e. information routinely available for many populations). The value of α will vary by age (Table 1) and, to a lesser extent, period. We applied the adjustment first by allowing α to vary over both age and period and secondly using a simpler approach where we assumed α did not vary by period.

a. Adjustment for cancer patient mortality using cause of death statistics in 1987–2002 aggregated in five-year periods and five-year age groups.
b. Adjustment for cancer patient mortality using cause of death statistics in 2002 aggregated in five-year age groups.

The formula for the adjustment provides insight into when we might expect to see differences between the standard method (method 2) and the theoretically correct method. It is clear that the bias resulting from using the standard method increases with increasing α but less obvious that the bias also depends on the magnitude of the underlying expected survival probability. For example, the expected 10-year survival probability for a 25 year old Swedish female in 2008 was 0.997. Even if all deaths among 25–34 year olds were due to cancer (making the theoretically correct expected survival 1.000) the bias in expected survival resulting from using the standard method will be 0.003. Fig. 1 shows the 10-year expected survival for Swedish females in 2008 together with the adjusted 10-year expected survival for various hypothetical values of α (the proportion of deaths due to cancer). We see that expected survival is sufficiently high for ages up to 55 years that even if a very large proportion of deaths in the general population are due to cancer the bias that results from using the standard method will be negligible. For persons aged 75 years and older, where expected survival is lower, the proportion of deaths due to a specific cancer rarely exceeds 5% (Table 1). For example, the proportion of deaths due to breast cancer among women aged 75–84 is 2.8% whereas it is greater than 10% for women aged 35–64. The exception is prostate cancer, which accounts for 7.2% of deaths among men aged 75–84.

We calculated cohort-based cumulative RSRs for patients diagnosed during the year 1987 and the periods 1987–2001, 1987–1991, 1992–1996, and 1997–2001. Patients were followed until the end of 2002 or a maximum of 15 years after diagnosis. The RSRs were calculated for males and females as well as for both sexes combined and for different age groups. Of these, the results for patients diagnosed 1987, the age groups 0–54, 55–74, 75–w, and 0–w, and both sexes combined are reported here. When expected survival is high (i.e. among younger persons) the bias introduced into the RSRs will be small even if the proportion of cancer deaths is high for the particular age group. There is as such no need to specifically analyse younger age groups.

We used a publicly available SAS macro[7,8] that implements the Hakulinen method[9] to perform the analysis. The macro was updated to incorporate the use of annual expected survival probabilities[10] and further adapted by the first author

Table 1 – Percentage of deaths due to cancer in the USA in 2007 (only for all sites) and Sweden in 2008 (all sites and site-specific). All sites combined, colorectal, lung, and malignant skin cancer for males and females, and breast cancer for females and prostate cancer for males.

Age	All sites (USA)	All sites	Colorectal	Lung	Skin	Breast	Prostate
0–w	23.2	23.8	2.9	3.9	0.6	3.2	5.6
<1	0.2	0.4	0.0	0.0	0.0	0.0	0.0
1–4	7.7	17.2	0.0	0.0	0.0	0.0	0.0
5–14	15.7	19.3	0.0	0.0	0.0	0.0	0.0
15–24	4.9	8.5	0.0	0.0	0.7	0.0	0.0
25–34	8.1	15.6	1.5	0.6	1.0	6.0	0.0
35–44	16.7	29.6	3.4	2.5	2.6	11.6	0.1
45–54	27.2	38.2	3.7	6.0	1.7	15.0	1.1
55–64	35.9	42.9	4.5	9.9	1.3	10.0	3.0
65–74	35.6	41.5	5.1	8.7	1.0	6.3	5.4
75–84	25.1	26.1	3.4	4.2	0.5	2.8	7.2
85–94	12.3[a]	13.0	1.8	1.1	0.3	1.4	6.4
95+		4.6	0.7	0.1	0.2	0.9	3.7

Data for USA extracted from Table 9 of National Vital Statistics Reports, volume 58, number 19 (http://www.cdc.gov/nchs/data/nvsr/nvsr58/nvsr58_19.pdf).

[a] Figure for USA is for ages 85+.

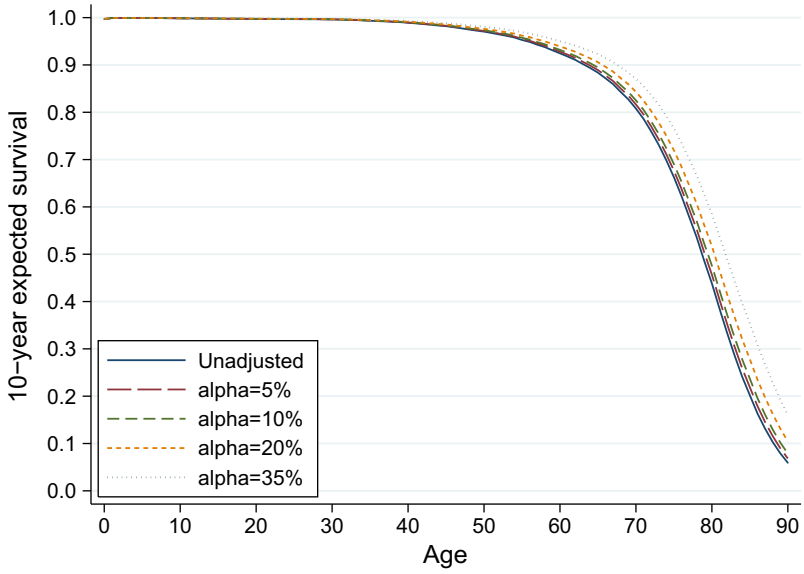

Fig. 1 – 10-year expected survival for Swedish females in 2008 together with the adjusted 10-year expected survival for a cancer free population under the assumption that 5%, 10%, 20%, and 35% of deaths, respectively, are due to cancer.

of this article to facilitate easy use of different expected survival probabilities.

3. Results

As expected, the RSRs calculated with expected survival probabilities from the general population (method 2) are overestimated compared to the RSRs calculated with expected survival probabilities using the reference method (method 1). The differences (i.e. bias) increase with both length of follow-up and age at diagnosis (Table 2).

For all sites combined, the RSRs are overestimated between 3.3 and 4.5 percent units after 10 years of follow-up for all ages combined and for the older age groups (Table 2). For colorectal cancer the bias is between 0.5 and 0.8 percent unit. For lung and skin cancer the bias is small for all age groups due to low prevalence, and low excess mortality, respectively. For breast cancer the bias is between 0.6 and 1.2 percent units and for prostate cancer between 1.2 and 2.6 percent units.

We also see from Table 2 that the simple method of adjustment (methods 3a and 3b) performs well in reducing the bias. For all sites combined, the bias decreased to less than 0.1 percent units in all but the oldest age group using the finer adjustment method (method 3a). For colorectal and lung cancer the bias was eliminated. For skin and breast cancer a small overestimation remains for the 75+ age group of 0.2 and 0.4, respectively, and for prostate cancer the overestimation of the 10-year RSR by 2.6 percent units for the 75+ age group was eliminated.

To save space we reported results for both sexes combined except for breast and prostate cancer. However, the differences between the methods 1 and 2 were, in general, slightly larger for males than females. The results for the periods not reported in the article were very similar to the results for patients diagnosed in 1987.

4. Discussion

Our evaluation of the bias introduced into the RSR by using expected survival probabilities from the general population shows that for most cancer types the bias will be sufficiently small that it can be ignored in practical applications. This is especially true when prevalence is low (i.e. rare cancers and younger age groups). However, for common cancer types, for older age groups, and for all cancers combined our results show that the bias in the RSR can be up to five percent units after 10 years of follow-up.

There is no general definition of what constitutes a large bias, but some indication of the relative size of this bias can be gained by comparing it to the width of the 95 percent confidence limit (CL) of the RSR, although the bias is of course not a measure of random variation. For all sites combined, the bias is approximately the same size as the CL width after one year of follow-up for all ages combined and for the two older age groups. After 10 years the bias has increased to approximately four times the size of the CL width for all ages combined and the 55–74 age group, and to three times the size of the CL for the 75+ age group. For colorectal cancer the bias is below 20% of the CL after 10 years of follow-up, and for lung and skin cancer the bias is negligible compared to the CL. For breast cancer the bias is approximately 30% of the CL after 10 years of follow-up for all ages combined and the 55–74 age group, and to 15% of the CL for the 75+ age group. For prostate cancer the bias is 60% of the CL after 10 years for all ages combined, 40% for the 55–74 age group, and 50% for the 75+ age group.

For prostate cancer, the cruder adjustment (method 3b) causes the RSRs to be underestimated with 0.4 percent units after 10 years of follow-up, whereas for breast cancer both adjustments overestimates the RSRs with some 0.4 percent units after 10 years. This indicates that for prostate cancer the increased incidence (and subsequent prevalence) during

Table 2 – Cohort-based cumulative relative survival ratios (RSR) for patients diagnosed in 1987, calculated with expected survival probabilities estimated from individual population records where cancer patients were censored at the time of diagnosis (Method 1, reference) and the difference in percent units to RSRs calculated with unadjusted (Method 2) and adjusted (Method 3a and 3b) expected survival probabilities from the general population. Positive differences indicate an overestimation of the RSRs compared to the reference and negative differences indicate an under estimation.

Age	Method of estimation	Males and females Years since diagnosis					Males and females Years since diagnosis				
		1	2	5	10	15	1	2	5	10	15
All sites combined (ICD-10: C00-C97)							*Colorectal cancer (ICD-10: C18-C21)*				
0–w	RSR (Method 1)	66.7	58.2	47.6	40.8	38.7	70.0	60.5	47.9	42.4	40.9
	Method 2	0.6	1.0	2.0	3.3	4.3	0.1	0.2	0.3	0.5	0.6
	Method 3a	0.0	0.1	0.1	0.1	–0.1	0.0	0.0	0.0	0.0	–0.0
	Method 3b	–0.1	–0.1	–0.3	–0.4	–0.5	–0.0	–0.0	–0.0	–0.0	–0.1
0–54	RSR (Method 1)	84.2	77.0	67.2	60.7	57.2	81.4	72.1	60.9	56.2	53.4
	Method 2	0.1	0.1	0.3	0.8	1.4	0.0	0.0	0.0	0.1	0.1
	Method 3a	0.0	0.0	0.0	0.0	0.0	–0.0	–0.0	–0.0	–0.0	–0.0
	Method 3b	0.0	0.0	–0.0	0.0	0.0	–0.0	–0.0	–0.0	–0.0	–0.0
55–74	RSR (Method 1)	68.4	58.7	46.6	37.9	33.3	75.7	65.0	49.2	42.1	40.4
	Method 2	0.5	0.8	1.8	3.3	4.7	0.1	0.1	0.2	0.4	0.6
	Method 3a	0.0	0.1	0.1	0.1	–0.0	0.0	0.0	0.0	0.0	–0.0
	Method 3b	–0.1	–0.1	–0.2	–0.3	–0.5	–0.0	–0.0	–0.0	–0.1	–0.1
75–w	RSR (Method 1)	55.7	47.7	36.4	26.8	22.1	61.0	52.3	42.2	36.7	31.5
	Method 2	0.9	1.6	3.0	4.5	5.5	0.1	0.2	0.5	0.8	1.0
	Method 3a	0.1	0.1	0.1	–0.2	–0.7	0.0	0.0	0.0	0.0	–0.0
	Method 3b	–0.1	–0.2	–0.4	–0.8	–1.2	0.0	–0.0	–0.0	–0.0	–0.1
Lung cancer (ICD-10: C34)							*Malignant skin cancer (ICD-10: C43-C44)*				
0–w	RSR (Method 1)	27.2	13.5	8.0	5.6	4.8	95.4	91.2	84.3	77.9	77.1
	Method 2	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.1	0.1	0.2
	Method 3a	0.0	–0.0	–0.0	–0.0	–0.0	0.0	0.0	0.0	0.1	0.1
	Method 3b	–0.0	–0.0	–0.0	–0.0	–0.0	0.0	0.0	0.0	0.0	0.0
0–54	RSR (Method 1)	37.7	20.4	10.7	8.7	7.0	97.0	93.4	87.4	83.0	80.0
	Method 2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Method 3a	–0.0	–0.0	–0.0	–0.0	–0.0	0.0	0.0	0.0	0.0	–0.0
	Method 3b	–0.0	–0.0	–0.0	–0.0	–0.0	0.0	0.0	–0.0	–0.0	–0.0
55–74	RSR (Method 1)	30.7	15.2	9.0	5.6	4.5	95.7	91.2	85.8	80.3	80.7
	Method 2	0.0	0.0	0.1	0.1	0.1	0.0	0.0	0.1	0.2	0.3
	Method 3a	0.0	–0.0	–0.0	–0.0	–0.0	0.0	0.0	0.0	0.1	0.1
	Method 3b	–0.0	–0.0	–0.0	–0.0	–0.0	0.0	0.0	–0.0	0.0	0.0
75–w	RSR (Method 1)	15.8	6.8	3.9	2.6	2.4	94.2	90.1	80.8	68.1	67.2
	Method 2	0.0	0.0	0.0	0.0	0.1	0.0	0.1	0.2	0.3	0.4
	Method 3a	–0.0	–0.0	–0.0	–0.0	–0.0	0.0	0.0	0.1	0.2	0.2
	Method 3b	–0.0	–0.0	–0.0	–0.0	–0.0	0.0	0.0	0.0	0.1	0.1
Breast cancer (ICD-10: C50)		Females					*Prostate cancer (ICD-10:C61)*				
0–w	RSR (Method 1)	94.7	90.5	78.8	68.2	62.1	85.8	77.7	59.5	42.6	31.9
	Method 2	0.1	0.2	0.4	0.6	0.8	0.3	0.5	1.0	1.5	1.7
	Method 3a	0.0	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	–0.0
	Method 3b	0.0	0.0	0.1	0.1	0.1	–0.0	–0.1	–0.1	–0.1	–0.1
0–54	RSR (Method 1)	97.4	93.5	80.5	69.2	63.6	87.6	86.4	48.3	38.2	29.6
	Method 2	0.0	0.1	0.1	0.3	0.4	–0.0	–0.0	0.0	0.0	0.1
	Method 3a	0.0	0.0	0.0	0.0	0.0	–0.0	–0.0	–0.0	–0.0	–0.0
	Method 3b	0.0	0.0	0.0	0.0	0.0	–0.0	–0.0	–0.0	–0.0	–0.0
55–74	RSR (Method 1)	95.3	91.1	80.9	71.5	64.9	89.9	81.3	63.1	45.1	33.4
	Method 2	0.1	0.2	0.4	0.7	0.9	0.2	0.3	0.6	1.2	1.6
	Method 3a	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0
	Method 3b	0.0	0.0	0.0	0.1	0.1	0.0	–0.0	–0.0	0.0	–0.0

Table 2 – (*continued*)

Age	Method of estimation	Males and females Years since diagnosis					Males and females Years since diagnosis				
		1	2	5	10	15	1	2	5	10	15
75–w	RSR (Method 1)	90.8	85.9	71.6	54.6	41.4	81.4	73.5	55.2	38.6	29.7
	Method 2	0.2	0.3	0.7	1.2	1.4	0.4	0.8	1.7	2.6	3.3
	Method 3a	0.1	0.1	0.2	0.4	0.3	0.0	0.1	0.1	0.0	–0.4
	Method 3b	0.0	0.1	0.2	0.3	0.3	–0.1	–0.1	–0.3	–0.4	–0.6

Method 1: Expected probabilities estimated from the database with individual population records where the cancer patients were censored at their time of diagnosis.
Method 2: Expected probabilities estimated from aggregated official population and mortality statistics unadjusted for cancer patient mortality.
Method 3a: Method 2 adjusted for cancer patient mortality using cause of death statistics in 1987-2002 aggregated in five-year periods and five-year age groups.
Method 3b: Method 2 adjusted for cancer patient mortality using cause of death statistics in 2002 aggregated in five-year age groups.

the late 1990s and early 2000s[11] has caused the cruder adjustment, using only cause of death statistics from 2002, to overestimate the decease-free expected survival probabilities.

For breast cancer, where a nationwide well-established screening programme[12] has been in place since the late 1980s, early 1990s, it is likely that the number of screen-detected cancers has not varied as much during the period.

In the analysis we increased the prevalence in the population for some of the cancer types by combining colon and rectal cancer, and melanoma and other malignant cancers of the skin. Despite this the bias remained small and can be expected to be even smaller when these cancers are analysed separately.

Ederer et al. concluded 50 years ago that, since we are usually analysing patients with specific forms of cancer, we do not need to make an adjustment when estimating expected survival from general-population life tables. However, they made this statement on the basis of analysing lung cancer and attributed the small differences between the RSRs calculated with and without adjustment for smoking habits to the low survival of the lung cancer patients, i.e. low prevalence. This statement is consistent with our results, but it was made at a time when the survival of lung cancer patients was much lower than it is today. Our results also show that the findings for lung cancer (where we also found only a small bias of no more than 0.1 percent units) do not generalise to other cancer types (where the bias can be as much as 3.3 percent units) and definitely not to all sites combined (where the bias can be over 5 percent units).

The prevalence of some cancers has increased significantly during recent decades due to a combination of increasing incidence and survival (due to screening and improved diagnostic methods and treatment). The incidence and prevalence of prostate cancer have increased the most in Sweden[11] during the past decade. Oksanen[5] concluded that the effect of excluding prevalent prostate cancer cases from the population in order to form a disease-free population had only minor effect on the mortality estimates, even though it is the most common cancer among men. However, she analysed patients diagnosed 1970–1993 when prostate cancer incidence was considerably lower than it is today.

Relative survival is defined as the observed survival of the cancer patients divided by the expected survival of a comparable group from the general population, free from the cancer under study. This paper addresses only the potential bias associated with assuming the general population is 'free from the cancer under study' and the proposed adjustment corrects only for this one specific potential bias. We must assess separately whether it is reasonable to assume that the patients are 'comparable' to the general population (after excluding individuals with the specific cancer under study) with respect to mortality other than cancer. It is known, for example, that this assumption is not strictly true for lung cancer, since the patients experience excess cardiovascular disease mortality compared to the general population. Neither is this assumption true if cancer incidence depends on social class (such as for breast cancer) but this can be accounted for by stratifying the general population life tables by social class. In addition, there are other assumptions that must be considered when estimating and interpreting relative survival[13], such as the independence of mortality due to the specific cancer and mortality due to other causes that are not addressed in this paper.

Although our paper addresses only one of several potential biases, it is a bias most practitioners are well aware of but one that is very difficult to study empirically. Taking advantage of the population registers in Sweden we have been able to demonstrate that the bias is, for most practical applications, ignorable but that there exist some applications (prostate cancer and 'all sites') where the bias is sufficiently large that we would recommend using the adjustment described in this paper.

Novelty and impact

To our knowledge, this is the first systematic evaluation ever made of the bias introduced into the relative survival ratios by including cancer patient mortality in the expected survival probabilities. In Sweden we have the unique opportunity to evaluate the size of this bias due to our computerised population registers that cover the entire nation.

Conflict of interest statement

None declared.

Sources of support

We thank the Swedish Cancer Society (Cancerfonden) for financial support.

Appendix A

Adjusting expected survival probabilities from the general population for cancer patient mortality. Only information regarding the numbers and causes of death (e.g. from cause of death statistics) is needed to do the adjustment. Survival probabilities were estimated for sex, age, and period (year) specific groups. Subscripts are omitted in the equations in order to simplify the expressions. Adapted from Dickman et al., appendix 1.[14]

$$P_{\bar{c}} = P_{gp}^{(1-\alpha)}, \quad (1)$$

where, $P_{\bar{c}}$ = Probability of surviving at least one additional period (year), for individuals not diagnosed with the specific cancer under study. P_{gp} = Probability of surviving at least one additional period (year) in the general population, e.g. expected survival probabilities from ordinary life tables. $\alpha = \frac{D_c}{D_T}$, Proportion of deaths in the general population due to the specific cancer under study D_T = Total number of deaths in the general population from cause of death statistics, in a relevant period and age group. D_c = Number of deaths in the general population due to the specific cancer under study from cause of death statistics, in a relevant period and age group.

For all sites combined and for each of the five individual cancer types the expected survival probabilities was adjusted for cancer patient mortality; (3a) using cause of death mortality statistics from 1987–2002 aggregated in five-year periods and five-year age groups, (3b) using cause of death statistics from 2002 aggregated in five-year age groups and applied to all years 1987–2002. The adjustments were made separately for males and females.

REFERENCES

1. Berrino F, De Angelis R, Sant M, et al. Survival for eight major cancers and all cancers combined for European adults diagnosed in 1995–1999: results of the EUROCARE-4 study. *Lancet Oncol* 2007;**8**(9):773–83.
2. Coleman M, Quaresma M, Berrino F, et al. Cancer survival in five continents: a worldwide population-based study (CONCORD). *Lancet Oncol* 2008;**9**(8):730–56.
3. Coleman MP, Forman D, Bryant H, et al. Cancer survival in Australia, Canada, Denmark, Norway, Sweden, and the UK, 1995–2007 (the International Cancer Benchmarking Partnership): an analysis of population-based cancer registry data. *Lancet* 2011;**377**(9760):127–38.
4. Ederer F, Axtell LM, Cutler SJ. The Relative survival rate: a statistical methodology. *National Cancer Institute Monographs* 1961;**6**:101–21.
5. Oksanen H. Modelling the survival of prostate cancer patients (Doctoral dissertation). University of Tampere, Finland; 1998. p. 33 (5.5), 56 (6.2.2), 70 (7.5, 7.6), 73-74 (8.).
6. Statistical database [Internet database]. In: Statistics Sweden, Stockholm, Sweden. 2007. http://www.ssd.scb.se/databaser/makro/start.asp?lang=2.
7. Brenner H, Gefeller O, Hakulinen T. A computer program for period analysis of cancer patient survival. *Eur J Cancer* 2002;**38**(5):690–5.
8. Brenner H, Gefeller O, Hakulinen T, Arndt V. Period and periodh: period analysis of survival date. In: IMBE Statistical Software Archive Network: University of Erlangen, Nürnberg. http://www.imbe.med.uni-erlangen.de/cms/software_period.html.
9. Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 1982;**38**(4):933–42.
10. Arndt V, Talbäck M, Gefeller O, Hakulinen T, Brenner H. Modification of SAS macros for a more efficient analysis of relative survival rates. *Eur J Cancer* 2004;**40**(5):778–9.
11. Cancer Incidence in Sweden 2009. Stockholm, Sweden: National Board of Health and Welfare, Stockholm, Sweden; 2010.
12. Socialstyrelsens cancerrapport till regeringen – Lägesrapport (In English: The National Board of Health and Welfare's Cancer repport to the Government – Progress report). Stockholm, Sweden: National Board of Health and Welfare; 2006.
13. Hakulinen T, Seppä K, Lambert PC. Choosing the relative survival method for cancer survival estimation. Eur J Cancer 2011, doi:10.1016/j.ejca.2011.03.011.
14. Dickman P, Auvinen A, Voutilainen E, Hakulinen T. Measuring social class differences in cancer patient survival: is it necessary to control for social class differences in general population mortality? A Finnish population-based study. *J Epidemiol Community Health* 1998;**52**(11):727–34.