Cancer Epidemiology 36 (2012) 148–152

Contents lists available at SciVerse ScienceDirect

Cancer Epidemiology

The International Journal of Cancer Epidemiology, Detection, and Prevention

journal homepage: www.cancerepidemiology.net

Adjusting for the proportion of cancer deaths in the general population when using relative survival: A sensitivity analysis

Sally R. Hinchliffe [a,*], Paul W. Dickman [b], Paul C. Lambert [a,b]

[a] Department of Health Sciences, Centre for Biostatistics and Genetic Epidemiology, University of Leicester, Leicester, UK
[b] Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

ARTICLE INFO

Article history:
Received 9 August 2011
Accepted 20 September 2011
Available online 13 October 2011

Keywords:
Relative survival
Expected survival
Cancer
Sensitivity analysis
Life tables

ABSTRACT

Background: Relative survival is an extensively used method in population based cancer studies as it provides a measure of survival without the need for accurate cause of death information. It gives an estimate for the probability of dying from cancer in the absence of other causes by estimating the excess mortality in the study population when compared to an external group. The external group is usually the general population within a country or state and mortality estimates are taken from national life tables that are broken down by age, sex, calendar year and, where applicable, race/ethnicity. One potential bias when using relative survival that is most often overlooked occurs when there are a high proportion of deaths due to a specific cancer in the external group. *Methods*: This paper uses data from the Finnish Cancer Registry to illustrate, through the use of a simple sensitivity analysis, the impact that specific cancer deaths in the population mortality figures can have on the estimate of relative survival. *Results*: We found that when examining specific diseases such as breast cancer and colon cancer, the proportion of deaths due to these specific cancers in the general population is so small in comparison to the total mortality that they make little difference to the relative survival estimates. However, prostate cancer proved to be an exception to this. For all cancer sites combined the sensitivity analysis illustrates a major limitation for this type of analysis, particularly with the older age groups. *Conclusion*: We recommend that, with a classification of diseases as wide as all cancer sites, relative survival should not be used without appropriate adjustment.

1. Introduction

Relative survival is an extensively used method in population based cancer studies as it circumvents the need for accurate cause of death information. It does this by providing a measure of survival based on estimating the excess mortality in the study population when compared to an external group [1]. The external group is usually the general population within a country or state and mortality estimates are taken from national life tables stratified by age, sex, calendar year and, where applicable, race or ethnicity. These mortality estimates are taken to be the mortality rates for cancer patients if they did not have cancer and so any excess mortality found in the cancer group is deemed to be due to cancer-related deaths [2]. However, in reality the population mortality estimates will also contain cancer deaths.

This study addresses one of the potential limitations when using relative survival that arises when there are a high proportion of deaths due to a specific cancer in the external group. If this is the case then the excess mortality will most likely be under-estimated, producing over-estimates of relative survival in the cancer group. When faced with this limitation most papers will cite Ederer et al. [3], who discussed that it was reasonable to assume that the proportion of deaths due to a specific disease was negligible in comparison to the total mortality in the general population [3]. This assumption is questionable for more common cancers, particularly in the older age groups.

This paper aims to show how a simple sensitivity analysis can be performed to assess the impact that specific cancer deaths in the population mortality figures can have on the estimate of relative survival. We study the effects by age group and for selected cancer sites.

2. Methods

2.1. Relative survival

Relative survival is the ratio of the observed survival in the study population to the expected survival in a comparable external

* Corresponding author.
E-mail address: srh20@leicester.ac.uk (S.R. Hinchliffe).

1877-7821/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.canep.2011.09.007

group [3]. It can be written as:

$$R(t) = \frac{S(t)}{S^*(t)} \quad (1)$$

where $S(t)$ is the observed survival, $S^*(t)$ is the expected survival and t is the time from diagnosis [4].

Expected survival is calculated using life tables that match the study population to the external population usually by age, sex and calendar year [2]. Several methods have been developed to estimate the expected survival, including the Ederer I and II [3,5] and the Hakulinen methods [6]. Based on recent recommendation, the Ederer II method is used for all analyses in this paper [7,8]. This method considers the matched individuals to be at risk until the corresponding cancer patient dies or is censored. Relative survival is often calculated in age groups, which are sometimes pooled to obtain an age standardised estimate.

2.2. Sensitivity analysis

Data on breast cancer (ICD-0-3: C50O-C509), colon cancer (ICD-0-3: C18O-C189, C260), prostate cancer (ICD-0-3: C619) and all cancer sites combined (ICD-0-3: COOO-C809) were used in the analyses. These data were obtained from the Finnish Cancer Registry for patients diagnosed in the years 1995 to 2007 inclusive. The population mortality data were obtained from the Human Mortality Database [9]. Only adults were considered in our analyses and anyone diagnosed through autopsy was excluded. All relative survival analyses were carried out separately for the age groups 18–44, 45–59, 60–74, 75–84 and 85+. In order to obtain up-to-date estimates of 10 year relative survival a period analysis approach was adopted. The relative survival estimates were derived from data on the survival experience of patients in the 2005–2007 period [10].

An initial relative survival analysis was carried out using the unadjusted population mortality data. The population mortality data were then modified to illustrate three alternative scenarios. This was done by denoting the probability of dying in the external group as q, the probability of dying from the cancer of interest in the external group as q_c and the probability of dying from other causes in the external group as q_o such that

$$q = q_o + q_c \quad (2)$$

It should be noted that q, q_o and q_c are yearly probabilities which will vary by age, sex and calendar year. In relative survival we assume q_c to be a very small proportion of q and so ignore any bias that may result in just using q. However, we want to use the probability of dying from other causes in the external group (q_o) rather than the probability of dying from either cancer or other causes in the external group (q) to see what influence the proportion of deaths due to the cancer of interest has. If we let $\alpha = q_c/q$ denote the proportion of deaths in the external group due to the cancer of interest then we can calculate q_o as follows:

$$q_o = q(1 - \alpha) \quad (3)$$

This adjustment was applied assuming that 2%, 5% or 10% of the deaths in the external group were due to the cancer of interest (i.e. $\alpha = 0.02$, 0.05 and 0.1). Writing this adjustment in terms of the expected survival, $p^* = 1 - q$, the adjusted expected survival, p_o^*, can be written as

$$p_o^* = p^* + \alpha(1 - p^*) \quad (4)$$

This simple sensitivity analysis was carried out for data on breast, colon and prostate cancer for each age group. Analyses were carried out on females for breast cancer, males for prostate cancer and both males and females combined for colon cancer. It is common for relative survival analyses to be carried out on all cancer sites combined [11–14]. This is usually done to obtain a single summary measure showing overall trends of cancer survival over time. These estimates are often used as a "surveillance tool" in policy making [15]. An analysis on all cancer sites combined was therefore carried out, for which additional adjustments of 20% and 30% were made (i.e. $\alpha = 0.2$ and 0.3).

3. Results

Table 1 gives approximate percentages of deaths across age groups due to each of the 3 specific cancer sites, as well as that of all cancer sites, within the population of Finland during the year 2000. Each age group specific percentage was calculated by dividing the number of deaths due to the cancer of interest (obtained from the Finnish cancer registry) by the total number of deaths for that age group (obtained from the Human Mortality Database [9]). The proportions of deaths due to cancer are highest in the 60–74 age group for colon cancer, prostate cancer and all cancer sites combined. For breast cancer, the highest proportion of deaths due to cancer occurs in the 18–44 age group. Due to increasing competing causes of death, the proportions of deaths due to cancer in the older age groups decrease, even though the total number of deaths increases with age.

Table 2 shows the expected survival for males aged 60 and 80 in the year 2000. The values for p^* give the unadjusted expected survival and p_2^*, p_5^* and p_{10}^* give the expected survival adjusted for 2%, 5% and 10% of deaths due to cancer. The numbers in bold give a clearer indication of how high proportions of cancer deaths have a greater impact on expected survival for older age groups, particularly those over the age of 80. Each adjustment leads to a higher expected survival. Although within each age group the absolute differences are fairly small, given that relative survival is a cumulative measure, these differences will accumulate over time. This is evident from the 5 year expected survival estimates given in the table. For example, for a patient aged 80 at diagnosis, the 5 year unadjusted expected survival is 0.5758 but when adjusted for 10% of deaths due to cancer the 5 year expected survival is 0.6092.

Relative survival curves for breast cancer, prostate cancer and all cancer sites combined are shown in Figs. 1–3, respectively. The corresponding figure for colon cancer can be found in the supplementary material. All four figures show that adjusting for a high proportion of deaths due to a particular disease makes little difference in the younger age groups. This is not surprising given that the probability of death due to other causes is very low in these groups. In fact, the relative survival curve for prostate cancer actually goes above 1 in the 18–44 age group, suggesting that this group has a better survival than the general population. For the older age groups the sensitivity analysis highlights some more noticeable differences. However, for breast cancer in particular, the proportions used to adjust the expected survival in the graph are relatively large in comparison to the estimated proportions given in Table 1. For example, in the 85+ age group it is estimated that only 0.4% of all deaths are due to breast cancer but the smallest adjustment made in the graph is 2%. The estimated biases associated with using uncorrected life tables are clearer to see in Table 3.

Table 1
Percentages of deaths in Finland in the year 2000 due to specific cancers.

Age	Breast	Colon	Prostate	All sites
18–44	13.3	0.4	0.1	15.9
45–59	12.4	1.7	1.5	29.2
60–74	4.8	2.0	4.3	32.9
75–84	1.5	1.3	3.3	18.0
85+	0.4	0.7	2.2	7.9

Table 2
Unadjusted and adjusted expected survival for males diagnosed in the year 2000.

ID	FU	Year	Age	p^*	p_2^*	p_5^*	p_{10}^*
1	1	2000	60	0.9873	0.9876	0.9879	0.9886
1	2	2001	61	0.9869	0.9872	0.9876	0.9882
1	3	2002	62	0.9850	0.9853	0.9857	0.9865
1	4	2003	63	0.9834	0.9837	0.9842	0.9850
1	5	2004	64	0.9826	0.9829	0.9834	0.9843
1	5 Year expected survival			0.9274	0.9288	0.9308	0.9344
2	1	2000	80	0.9161	0.9161	0.9187	0.9230
2	2	2001	81	0.9053	0.9072	0.9100	0.9148
2	3	2002	82	0.8962	0.8983	0.9014	0.9066
2	4	2003	83	0.8857	0.8880	0.8914	0.8971
2	5	2004	84	0.8746	0.8771	0.8808	0.8871
2	5 Year expected survival			0.5758	0.5815	0.5917	0.6092

Table 3 gives the percentage unit differences between the unadjusted 10 year relative survival estimates and the 10 year relative survival estimates adjusted for 2%, 5%, 10%, 20% and 30% of deaths due to cancer. It also includes a column showing the percentage unit differences for the estimated proportions of deaths due to cancer, (α), as given in Table 1. We see for breast cancer that the estimated bias associated with using unadjusted life tables does not exceed 0.5 percentage units for any age group when using the estimated values of α. For the youngest age group, where breast cancer comprises more than 10% of deaths in the population, the background mortality is so low that, even if we were to make no adjustment for these deaths, the resulting bias on the relative survival estimates would be relatively minor. For the oldest age group, on the other hand, breast cancer comprises such a small proportion of the total number of deaths (0.4%, Table 1) that the resulting bias on the relative survival estimates is also relatively minor. This result can also be seen for colon cancer.

A similar result can be seen in the younger age groups for prostate cancer. However, in the older age groups (75+) the bias on the relative survival estimates is a potential cause for concern. In the oldest age group, where prostate cancer comprises 2.2% of all deaths in the population, the bias on the relative survival estimates is over 4 percentage units. The results for all cancer sites combined highlight a major limitation of this type of analysis when using unadjusted life tables. Biases are evident in the relative survival estimates for all age groups but particularly those over the age of 60. In the 85+ age group, a proportion of 7.9% of deaths in the population are estimated to be due to all cancers, which leads to a bias of over 10 percentage units in the relative survival estimates.

4. Discussion

This paper has shown a simple method that can be used to assess the potential impact of using regional or national life tables for the probability of death due to other causes. It has also provided an equation (see Eq. (4)) that can be used in future analyses to adjust population mortality data sets in a similar way.

The above sensitivity analysis demonstrates that, when examining specific diseases such as colon cancer or breast cancer using relative survival, the assumption made by Ederer et al. in 1961 is usually reasonable. The proportion of deaths due to specific cancers in the general population is small in comparison to the total mortality. In the younger age groups these proportions are of little importance as the probability of dying is low anyway. In the older age groups the proportion of deaths due to specific cancers would need to reach at least 2% before important differences occur. Specific cancers rarely reach such proportions in the older age groups due to the large amount of competing causes of death. However, prostate cancer has proven to be an exception to this in

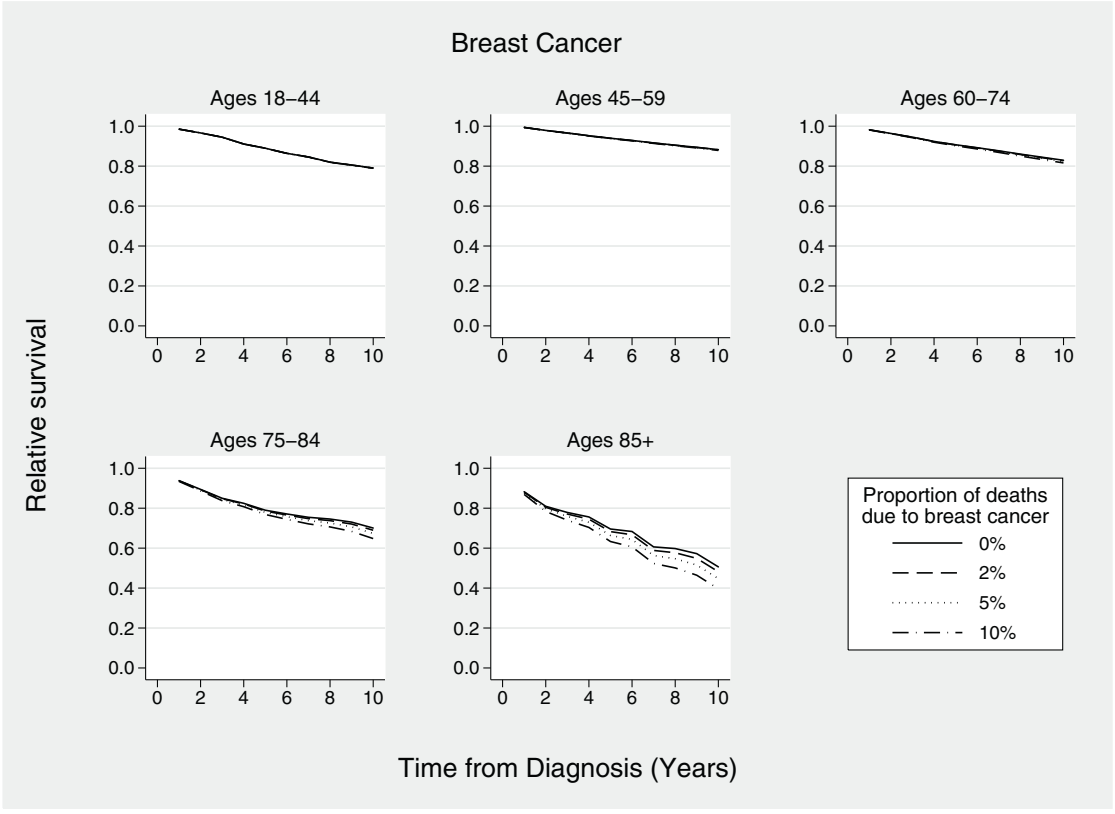

Fig. 1. Adjusting for varying proportions of breast cancer deaths in the general population.

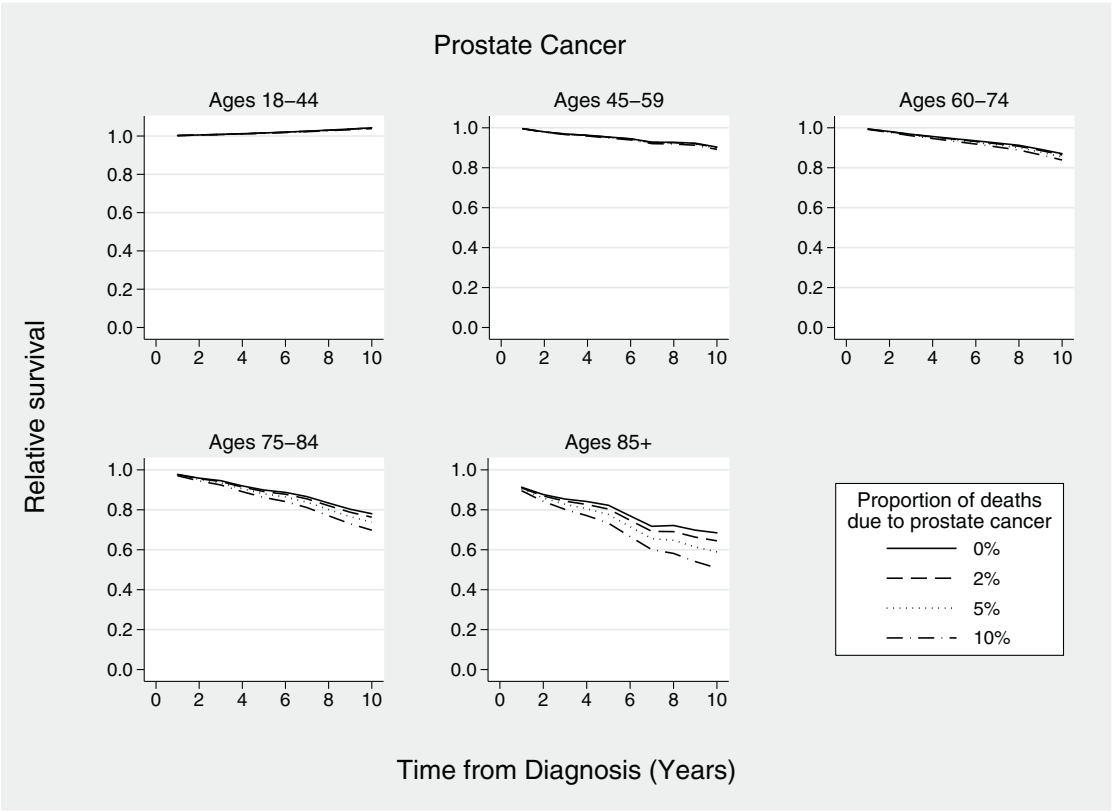

Fig. 2. Adjusting for varying proportions of prostate cancer deaths in the general population.

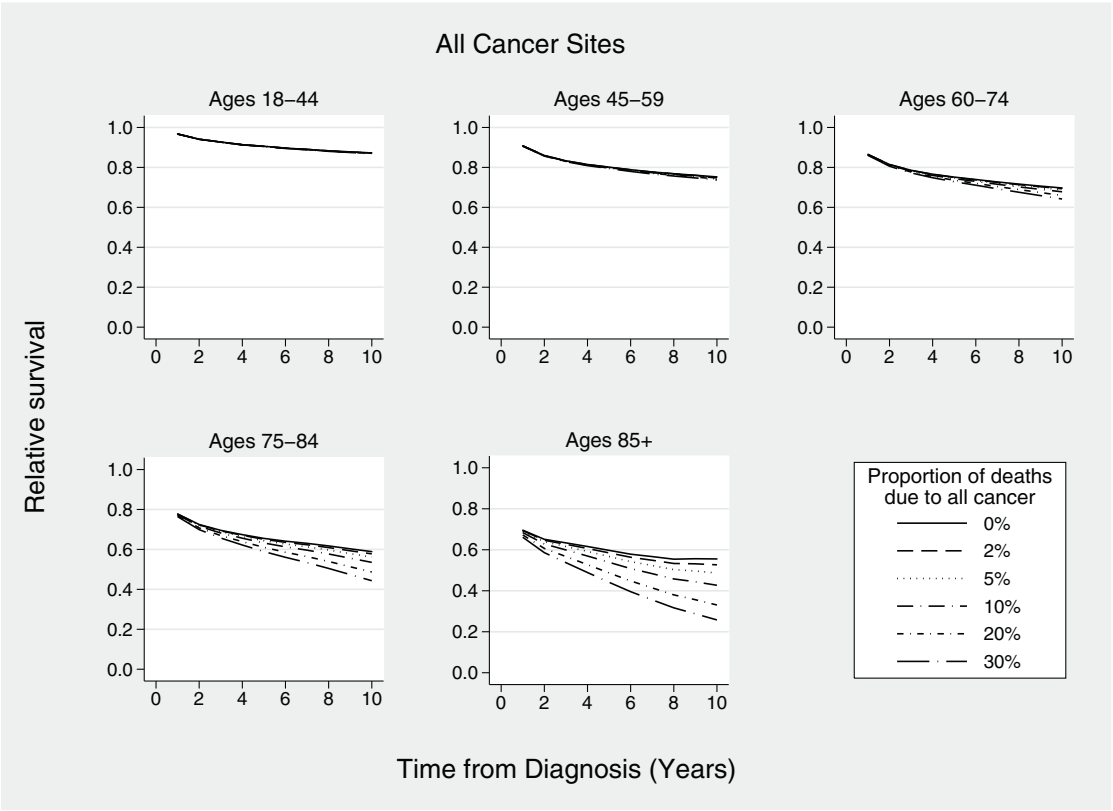

Fig. 3. Adjusting for varying proportions of all cancer deaths in the general population.

Table 3
Percentage unit differences in 10 year relative survival estimates between values with no adjustment (i.e. 0%) and adjusted values (i.e. 2%, 5%, 10%, 20%, and 30% and estimates α from Table 1).

Site	Age	2%	5%	10%	20%	30%	Estimated
Breast	18–44	0.02	0.06	0.12	–	–	0.16
	45–59	0.08	0.19	0.39	–	–	0.36
	60–74	0.27	0.67	1.34	–	–	0.43
	75–84	1.09	2.70	5.28	–	–	0.49
	85+	2.44	5.86	10.97	–	–	0.50
Colon	18–44	0.02	0.06	0.12	–	–	0.01
	45–59	0.09	0.22	0.44	–	–	0.08
	60–74	0.34	0.85	1.69	–	–	0.28
	75–84	0.96	2.36	4.59	–	–	0.47
	85+	2.85	6.84	12.80	–	–	1.01
Prostate	18–44	0.09	0.22	0.44	–	–	0.05
	45–59	0.23	0.57	1.14	–	–	0.39
	60–74	0.65	1.60	3.17	–	–	1.21
	75–84	1.73	4.26	8.26	–	–	2.41
	85+	4.01	9.54	17.63	–	–	4.39
All	18–44	0.03	0.06	0.13	0.25	0.38	0.29
	45–59	0.10	0.25	0.51	1.01	1.51	1.56
	60–74	0.38	0.96	1.90	3.73	5.52	4.65
	75–84	1.12	2.76	5.38	10.21	14.57	6.78
	85+	2.88	6.89	12.84	22.47	29.76	10.44

this paper, with an estimated bias of over 4 percentage units in the oldest age group.

The sensitivity analysis for all cancer sites combined illustrates a major limitation for this type of analysis, particularly with the older age groups. The proportion of deaths due to all cancers combined is reasonably high in these age groups, meaning that, when not adjusted for, the relative survival estimates are over-estimated by as much as 9 percentage units. It is therefore advised that, if relative survival is used for all cancer sites combined, an adjustment be made to the probability of dying in the external group as demonstrated in this paper. With a classification of diseases as wide as all cancer sites some care has to be taken when using relative survival. Although we do not advocate analyses using all cancer sites, relative survival is often used with such data in order to show overall cancer trends, usually on fact sheets made available to patients and the public. This is ill-advised as, although carrying out such analyses may be fairly straightforward, interpreting the estimates is near impossible.

The percentages in Table 1 may be unreliable for the older age groups as the estimates are based on cause of death information. These figures are most likely to be over-estimates as in the elderly population it is thought that when cancer is believed to be present before death it is usually certified as the cause of death even if it is not [16].

The method described in this paper only makes adjustments for the assumption that the proportion of deaths due to a specific disease is negligible in comparison to the total mortality in the general population as discussed by Ederer et al. in 1961. Other assumptions, such as that of independence between the mortality associated with the disease of interest and the mortality associated with other causes, are presumed to be reasonable.

Conflict of interest statement

None of the authors in this paper have any conflicts of interest to declare.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.canep.2011.09.007.

References

[1] Dickman PW, Adami HO. Interpreting trends in cancer patient survival. J Intern Med 2006;260:103–17.
[2] Sarfati D, Blakely T, Pearce N. Measuring cancer survival in populations: relative survival vs cancer-specific survival. Int J Epidemiol 2010;39:598–610.
[3] Ederer F, Axtell L, Cutler S. The relative survival rate: a statistical methodology. Natl Cancer Inst Monogr 1961;6:101–21.
[4] Lambert PC, Dickman PW, Nelson CP, Royston P. Estimating the crude probability of death due to cancer and other causes using relative survival models. Stat Med 2010;29:885–95.
[5] Ederer F, Heise H. Instructions to IBM 650 programmers in processing survival computations. Methodological note No 10 end results evaluation section. Bethesda, MD: National Cancer Institute, 1959.
[6] Hakulinen T. Cancer survival corrected for heterogeneity in patient withdrawal. Biometrics 1982;38:933–42.
[7] Hakulinen T, Seppä K, Lambert P. Choosing the relative survival method for cancer survival estimation. Eur J Cancer 2011;47:2202–10.
[8] Rutherford MR, Dickman PW, Lambert PC. Comparison of methods for calculating relative survival in population-based studies. Cancer Epidemiology (in press).
[9] Human Mortality Database (2008). The Human Mortality Database [electronic resource]. Berkley: University of California and Rostock: Max Planck Institute for Demographic Research. [http://www.mortality.org/].
[10] Brenner H, Gefeller O. An alternative approach to monitoring cancer patient survival. Cancer 1996;78:2004–10.
[11] Myers MH, Hies LAG. Cancer patient survival rates: SEER program results for 10 years of follow-up. CA Cancer J Clin 1989;39:21–32.
[12] Jen-la A, Siegel R, Ward E, Murray T, Xu J, Smigal C, et al. Cancer statistics 2006. CA Cancer J Clin 2006;56:106–30.
[13] Brenner H. Long-term survival rates of cancer patients achieved by the end of the 20th century: a period analysis. The Lancet 2002;360:1131–5.
[14] Berrino F, Verdecchia A, Lutz JM, Lombardo C, Michell A, Capocaccia R. Comparative cancer survival information in Europe. Eur J Cancer 2009;45:901–8.
[15] Office for National Statistics. A cancer survival index for primary care trusts; 2010.
[16] Goldacre M. J. Cause-specific mortality: understanding uncertain tips of the disease iceberg. J Epidemiol Community Health 1993;47:491–6.